(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 December 2006 (14.12.2006)

PCT

(10) International Publication Number
WO 2006/131783 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IB2005/004037

(22) International Filing Date: 27 January 2005 (27.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/539,128	27 January 2004 (27.01.2004)	US
60/539,129	27 January 2004 (27.01.2004)	US
60/589,815	22 July 2004 (22.07.2004)	US
60/607,307	7 September 2004 (07.09.2004)	US
60/620,916	22 October 2004 (22.10.2004)	US
60/620,917	22 October 2004 (22.10.2004)	US
60/620,874	22 October 2004 (22.10.2004)	US
60/620,924	22 October 2004 (22.10.2004)	US
60/620,974	22 October 2004 (22.10.2004)	US
60/620,853	22 October 2004 (22.10.2004)	US
60/620,656	22 October 2004 (22.10.2004)	US
60/620,918	22 October 2004 (22.10.2004)	US
60/621,004	22 October 2004 (22.10.2004)	US
60/620,868	22 October 2004 (22.10.2004)	US
60/620,677	22 October 2004 (22.10.2004)	US
60/620,975	22 October 2004 (22.10.2004)	US
60/621,131	25 October 2004 (25.10.2004)	US
60/621,053	25 October 2004 (25.10.2004)	US
60/628,123	17 November 2004 (17.11.2004)	US
60/628,101	17 November 2004 (17.11.2004)	US
60/628,134	17 November 2004 (17.11.2004)	US
60/628,111	17 November 2004 (17.11.2004)	US
60/628,112	17 November 2004 (17.11.2004)	US
60/628,145	17 November 2004 (17.11.2004)	US
60/628,189	17 November 2004 (17.11.2004)	US
60/628,167	17 November 2004 (17.11.2004)	US
60/628,156	17 November 2004 (17.11.2004)	US
60/628,251	17 November 2004 (17.11.2004)	US
60/628,178	17 November 2004 (17.11.2004)	US
60/634,075	8 December 2004 (08.12.2004)	US
11/051,720	27 January 2005 (27.01.2005)	US

(71) Applicant *(for all designated States except US)*: **COMPUGEN USA, INC.**; 560 E. Winchester Blvd., Suite 500, San Jose, CA 96126 (US).

(72) Inventors: **POLLOCK, Sarah**; 16/2 Hoshea Street, 63506 Tel-Aviv (IL). **LEVINE, Zurit**; 4 Reihan Street, 46 419 Herzlia (IL). **NOVIK, Amit**; Hasheni Street, POB 1850, 40600 Beit-HaSharon (IL). **DAHARY, Dvir**; 36/4 Yehezkel Street, 62595 Tel-Aviv (IL). **SOREK, Rotem**; 15 Aharonovich Street, 76564 Rechovot (IL). **TOPORIK, Amir**; 20/2 Katzenelson Street, 58016 Azur (IL). **SAMEAH-GREENWALD, Shirley**; 33/7 Arlozorov Street, 44453 Kfar Saba (IL). **SELLA-TAVOR, Osnat**; 18 Kfar Kish, 19330 Israel (IL). **DIBER, Alexander**; 11/12 Grinspan Street, 75502 Rishon-LeZion (IL). **COJOCARU, Gad, S.**; 41/7 HaSayfan Street, 47248 Ramat-HaSharon (IL). **AYALON-SOFFER, Michal**; 14 HaNoter Street, 47210 Ramat-HaSharon (IL). **WALACH, Shira**; 40 Bnei Brith Street, 45265 Hod-HaSharon (IL). **AKIVA, Pinchas**; 8 HaMarganit Street, 52584 Ramat-Gan (IL). **KEREN, Naomi**; 9/31 HaOranim Street, 54052 Givat Shmuel (IL). **SHEMESH, Ronen**; 23/5 Hativat Givati Street, 71700 Modiln (IL).

(74) Agent: **BECKERS, Randall, J.**; Staas & Halsey LLP, 1201 New York Ave., NW, Suite 700, Washington, DC 20005 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS OF USE THEREOF FOR DIAGNOSIS OF LUNG CANCER

(57) Abstract: Novel markers for lung cancer that are both sensitive and accurate. These markers are overexpressed in lung cancer specifically, as opposed to normal lung tissue. The measurement of these markers, alone or in combination, in patient samples provides information that the diagnostician can correlate with a probable diagnosis of lung cancer. The markers of the present invention, alone or in combination, show a high degree of differential detection between lung cancer and non-cancerous states.

WO 2006/131783 A2

NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS OF USE THEREOF FOR DIAGNOSIS OF LUNG CANCER

FIELD OF THE INVENTION

The present invention is related to novel nucleotide and protein sequences that are diagnostic markers for lung cancer, and assays and methods of use thereof.

BACKGROUND OF THE INVENTION

Lung cancer is the primary cause of cancer death among both men and women in the U. S., with an estimated 172,000 new cases being reported in 1994. The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, only 16% of lung cancers are discovered before the disease has spread. Lung cancers are broadly classified into small cell or non-small cell lung cancers. Non-small cell lung cancers are further divided into adenocarcinomas, bronchoalveolar-alveolar, squamous cell and large cell carcinomas. Approximately, 75-85 percent of lung cancers are non-small cell cancers and 15-25 percent are small cell cancers of the lung.

Early detection is difficult since clinical symptoms are often not seen until the disease has reached an advanced stage. Currently, diagnosis is aided by the use of chest x-rays, analysis of the type of cells contained in sputum and fiberoptic examination of the bronchial passages. Treatment regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy.

Early detection of primary, metastatic, and recurrent disease can significantly impact the prognosis of individuals suffering from lung cancer. Non-small cell lung cancer diagnosed at an early stage has a significantly better outcome than that diagnosed at more advanced stages. Similarly, early diagnosis of small cell lung cancer potentially has a better prognosis.

Although current radiotherapeutic agents, chemotherapeutic agents and biological toxins are potent cytotoxins, they do not discriminate between normal and malignant cells, producing adverse effects and dose-limiting toxicities. There remains a need for lung cancer specific cancer markers. There remains a need for reagents and kits which can be used to detect the presence of lung cancer markers in samples from patients. There remains a need for methods of

screening and diagnosing individuals who have lung cancer and methods of monitoring response to treatment, disease progression and disease recurrence in patients diagnosed with lung cancer. There remains a need for reagents, kits and methods for determining the type of lung cancer that an individual who has lung cancer has. There remains a need for compositions which can specifically target lung cancer cells. There remains a need for imaging agents which can specifically bind to lung cancer cells. There remains a need for improved methods of imaging lung cancer cells. There remains a need for therapeutic agents which can specifically bind to lung cancer cells. There remains a need for improved methods of treating individuals who are suspected of suffering from lung cancer.

SUMMARY OF THE INVENTION

The background art does not teach or suggest markers for lung cancer that are sufficiently sensitive and/or accurate, alone or in combination.

The present invention overcomes these deficiencies of the background art by providing novel markers for lung cancer that are both sensitive and accurate. Furthermore, these markers are able to distinguish between different types of lung cancer, such as small cell or non-small cell lung cancer, and further between non-small cell lung cancer types, such as adenocarcinomas, squamous cell and large cell carcinomas. These markers are overexpressed in lung cancer specifically, as opposed to normal lung tissue. The measurement of these markers, alone or in combination, in patient (biological) samples provides information that the diagnostician can correlate with a probable diagnosis of lung cancer. The markers of the present invention, alone or in combination, show a high degree of differential detection between lung cancer and non-cancerous states.

According to preferred embodiments of the present invention, examples of suitable biological samples which may optionally be used with preferred embodiments of the present invention include but are not limited to blood, serum, plasma, blood cells, urine, sputum, saliva, stool, spinal fluid or CSF, lymph fluid, the external secretions of the skin, respiratory, intestinal, and genitourinary tracts, tears, milk, neuronal tissue, lung tissue, any human organ or tissue, including any tumor or normal tissue, any sample obtained by lavage (for example of the bronchial system or of the breast ductal system), and also samples of in vivo cell culture constituents. In a preferred embodiment, the biological sample comprises lung tissue and/or

sputum and/or a serum sample and/or a urine sample and/or any other tissue or liquid sample. The sample can optionally be diluted with a suitable eluant before contacting the sample to an antibody and/or performing any other diagnostic assay.

Information given in the text with regard to cellular localization was determined according to four different software programs: (i) tmhmm (from Center for Biological Sequence Analysis, Technical University of Denmark DTU, http://www.cbs.dtu.dk/services/TMHMM/TMHMM2.0b.guide.php) or (ii) tmpred (from EMBnet, maintained by the ISREC Bionformatics group and the LICR Information Technology Office, Ludwig Institute for Cancer Research, Swiss Institute of Bioinformatics, http://www.ch.embnet.org/software/TMPRED_form.html for transmembrane region prediction; (iii) signalp_hmm or (iv) signalp_nn (both from Center for Biological Sequence Analysis, Technical University of Denmark DTU, http://www.cbs.dtu.dk/services/SignalP/background/prediction.php) for signal peptide prediction. The terms "signalp_hmm" and "signalp_nn" refer to two modes of operation for the program SignalP: hmm refers to Hidden Markov Model, while nn refers to neural networks. Localization was also determined through manual inspection of known protein localization and/or gene structure, and the use of heuristics by the individual inventor. In some cases for the manual inspection of cellular localization prediction inventors used the ProLoc computational platform [Einat Hazkani-Covo, Erez Levanon, Galit Rotman, Dan Graur and Amit Novik; (2004) "Evolution of multicellularity in metazoa: comparative analysis of the subcellular localization of proteins in Saccharomyces, Drosophila and Caenorhabditis." Cell Biology International 2004;28(3):171-8.], which predicts protein localization based on various parameters including, protein domains (e.g., prediction of trans-membranous regions and localization thereof within the protein), pI, protein length, amino acid composition, homology to pre-annotated proteins, recognition of sequence patterns which direct the protein to a certain organelle (such as, nuclear localization signal, NLS, mitochondria localization signal), signal peptide and anchor modeling and using unique domains from Pfam that are specific to a single compartment.

Information is given in the text with regard to SNPs (single nucleotide polymorphisms). A description of the abbreviations is as follows. "T - > C", for example, means that the SNP results in a change at the position given in the table from T to C. Similarly, "M - > Q", for example, means that the SNP has caused a change in the corresponding amino acid sequence, from methionine (M) to glutamine (Q). If, in place of a letter at the right hand side for the nucleotide sequence SNP, there is a space, it indicates that a frameshift has occurred. A frameshift may also be indicated with a hyphen (-). A stop codon is indicated with an asterisk at the right hand side (*). As part of the description of an SNP, a comment may be found in parentheses after the above description of the SNP itself. This comment may include an FTId, which is an identifier to a SwissProt entry that was created with the indicated SNP. An FTId is a unique and stable feature identifier, which allows construction of links directly from position-specific annotation in the feature table to specialized protein-related databases. The FTId is always the last component of a feature in the description field, as follows: FTId=XXX_number, in which XXX is the 3-letter code for the specific feature key, separated by an underscore from a 6-digit number. In the table of the amino acid mutations of the wild type proteins of the selected splice variants of the invention, the header of the first column is "SNP position(s) on amino acid sequence", representing a position of a known mutation on amino acid sequence. SNPs may optionally be used as diagnostic markers according to the present invention, alone or in combination with one or more other SNPs and/or any other diagnostic marker. Preferred embodiments of the present invention comprise such SNPs, including but not limited to novel SNPs on the known (WT or wild type) protein sequences given below, as well as novel nucleic acid and/or amino acid sequences formed through such SNPs, and/or any SNP on a variant amino acid and/or nucleic acid sequence described herein.

Information given in the text with regard to the Homology to the known proteins was determined by Smith-Waterman version 5.1.2 using special (non default) parameters as follows:

-model=sw.model

-GAPEXT=0

-GAPOP=100.0

-MATRIX=blosum100

Information is given with regard to overexpression of a cluster in cancer based on ESTs. A key to the p values with regard to the analysis of such overexpression is as follows:

- library-based statistics: P-value without including the level of expression in cell-lines (P1)
- library based statistics: P-value including the level of expression in cell-lines (P2)
- EST clone statistics: P-value without including the level of expression in cell-lines (SP1)
- EST clone statistics: predicted overexpression ratio without including the level of expression in cell-lines (R3)
- EST clone statistics: P-value including the level of expression in cell-lines (SP2)
- EST clone statistics: predicted overexpression ratio including the level of expression in cell-lines (R4)

Library-based statistics refer to statistics over an entire library, while EST clone statistics refer to expression only for ESTs from a particular tissue or cancer.

Information is given with regard to overexpression of a cluster in cancer based on microarrays. As a microarray reference, in the specific segment paragraphs, the unabbreviated tissue name was used as the reference to the type of chip for which expression was measured. There are two types of microarray results: those from microarrays prepared according to a design by the present inventors, for which the microarray fabrication procedure is described in detail in Materials and Experimental Procedures section herein; and those results from microarrays using Affymetrix technology. As a microarray reference, in the specific segment paragraphs, the unabbreviated tissue name was used as the reference to the type of chip for which expression was measured. For microarrays prepared according to a design by the present inventors, the probe name begins with the name of the cluster (gene), followed by an identifying number. Oligonucleotide microarray results taken from Affymetrix data were from chips available from Affymetrix Inc, Santa Clara, CA, USA (see for example data regarding the Human Genome U133 (HG-U133) Set at www.affymetrix.com/products/arrays/specific/hgu133.affx; GeneChip Human Genome U133A 2.0 Array at www.affymetrix.com/products/arrays/specific/hgu133av2.affx; and Human Genome U133 Plus 2.0 Array at

www.affymetrix.com/products/arrays/specific/hgu133plus.affx). The probe names follow the Affymetrix naming convention. The data is available from NCBI Gene Expression Omnibus (see www.ncbi.nlm.nih.gov/projects/geo/ and Edgar et al, Nucleic Acids Research, 2002, Vol. 30, No. 1 207-210). The dataset (including results) is available from www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1133 for the Series GSE1133 database (published on March 2004); a reference to these results is as follows: Su et al (Proc Natl Acad Sci U S A. 2004 Apr 20;101(16):6062-7. Epub 2004 Apr 09). Probes designed by the present inventors are listed below.

>H61775_0_11_0
CCCCAGCTTTTATAGAGCGGCCCAAGGAAGAATATTTCCAAGAAGTAGGG
>M85491_0_0_25999
GACATCTTTGCATATCATGTCAGAGCTATAACATCATTGTGGAGAAGCTC
>M85491_0_14_0
GTCATGAAAATCAACACCGAGGTGCGGAGCTTCGGACCTGTGTCCCGCAG
>Z21368_0_0_61857
AGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAG
>HUMGRP5E_0_0_16630
GCTGATATGGAAGTTGGGGAATCTGAATTGCCAGAGAATCTTGGGAAGAG
>HUMGRP5E_0_2_0
TCTCATAGAAGCAAAGGAGAACAGAAACCACCAGCCACCTCAACCCAAGG
>D56406_0_5_0
TCTGACTTTTACGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGG
>F05068_0_0_5744
ACGGGAGGGAAGGAAGGTGTGCGGGAGGAGTTCTCTGTCTCCACTCCCCT
>F05068_0_0_5754
CAAGGGGAACTGACCGTTGGTCCCGAAGGTCTAGAAGTGAATGGGAGCAG
>F05068_0_8_0
CTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACT
>F05068_0_1_5751
TCTTAGCAGGTAGGTGCCGCAGACCCTGCGGGTTAAGAGGTGGGGTGGGG
>H38804_0_3_0

CGTAATTGCAGTGCATTTAGACAGGCATCTATTTGGACCTGTTTCTATCT
>HSENA78_0_1_0
TGAAGAGTGTGAGGAAAACCTATGTTTGCCGCTTAAGCTTTCAGCTCAGC
>R00299_0_8_0
CCAAGGCTCGTCTGCGCACCTTGTGTCTTGTAGGGTATGGTATGTGGGAC
>Z44808_0_8_0
AAAAGCATGAGTTTCTGACCAGCGTTCTGGACGCGCTGTCCACGGACATG
>Z44808_0_0_72347
ATGTTCTTAGGAGGCAAGCCAGGAGAAGCCGGGTCTGACTTTTCAGCTCA
>Z44808_0_0_72349
TCCTCCAGACCCAAAGCCACAACCCATCGCAAGTCAAGAACACTTTCCAG
>AA161187_0_0_433
ACCCTGGGTGGGCAAAAACGTGCTTTCCCGGACGGGGTTGAAGGGGAGAA
>AA161187_0_0_430
TGGAGACTGTTGCCCCACTCTGCAGATGCAGAAACGGAGGCTTGGCTGCT
>R66178_0_7_0
CCAGTGTGGTATCCTGGGAAACTCGGTTAAAAGGTGAGGCAGAGTACCAG
>HUMPHOSLIP_0_0_18458
AAGGAAGCAGGACCAGTGGATGTGAGGCGTGGTCGAAGAACAACAGAAAG
>HUMPHOSLIP_0_0_18487
ACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGG
>AI076020_0_3_0
ATCAGCACTGCCACCTACACCACGGTGCCGCGCGTGGCCTTCTACGCCGG
>T23580_0_0_902
GTGAAACCCCATTGGCTTCATTGGCTCCTTGATTTAAACCACGCCCGGCT
>T23580_0_0_901
TGAGTCCGTGTTATATCATCTGGTCTCATTGATAGGCGGGATAGGGAGGG
>M79217_0_9_0
TTTGTGGAATAGCAACCCATGGTTATGGCGAGTGACCCGACGTGATCTGG
>M62096_0_0_20588
AAGGCTTAGGTGCAAAGCCATTGGATACCATACCTGAGACCACACAGCCA

>M62096_0_7_0
ACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAGCAGAAA
>M78076_0_7_0
GAGAAGATGAACCCGCTGGAACAGTATGAGCGAAAGGTGAATGCGTCTGT
>T99080_0_0_58896
AACTCACAGCAAGAGCTGTGTTCCAGTTAGCTTTGCTACCAGTTATGCAG
>T08446_0_9_0
CATTTCCACTACGAGAACGTTGACTTTGGCCACATTCAGCTCCTGCTGTC
>HUMCA1XIA_0_0_14909
GCTGCAATCTAAGTTTCGGAATACTTATACCACTCCAGAAATAATCCTCG
>HUMCA1XIA_0_18_0
TTCAGAACTGTTAACATCGCTGACGGGAAGTGGCATCGGGTAGCAATCAG
>T11628_0_9_0
ACAAGATCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCATCATCCAG
>T11628_0_0_45174
TAAACAATCAAAGAGCATGTTGGCCTGGTCCTTTGCTAGGTACTGTAGAG
>T11628_0_0_45161
TGCCTCGCCACAATGGCACCTGCCCTAAAATAGCTTCCCATGTGAGGGCT
>HUMCEA_0_0_96
CAAGAGGGGTTTGGCTGAGACTTTAGGATTGTGATTCAGCTTAGAGGGAC
>HUMCEA_0_0_15183
CCTGGTGGGAGCCCATGAGAAGCGAGTTCTCTGTGCAACGGACTTAGTAA
>HUMCEA_0_0_15182
GCTCCCTGGAGCATCAGCATCATATTCTGGGGTGGAGTCTATCTGGTTCT
>HUMCEA_0_0_15168
TCCTGCCTGTCACCTGAAGTTCTAGATCATTCCCTGGACTCCACTCTATC
>HUMCEA_0_0_15180
TTTAACACAGGATTGGGACAGGATTCAGAGGGACACTGTGGCCCTTCTAC
>R35137_0_5_0
TATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAA
>Z25299_0_3_0

AACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTG

>HSSTROL3_0_0_12518

ATGAGAGTAACCTCACCCGTGCACTAGTTTACAGAGCATTCACTGCCCCA

>HSSTROL3_0_0_12517

CAGAGATGAGAGCCTGGAGCATTGCAGATGCCAGGGACTTCACAAATGAA

>HSS100PCB_0_0_12280

CTCAAAATGAAACTCCCTCTCGCAGAGCACAATTCCAATTCGCTCTAAAA

>R20779_0_0_30670

CCGCGTTGCTTCTAGAGGCTGAATGCCTTTCAAATGGAGAAGGCTTCCAT

The following list of abbreviations for tissues was used in the TAA histograms. The term "TAA" stands for "Tumor Associated Antigen", and the TAA histograms, given in the text, represent the cancerous tissue expression pattern as predicted by the biomarkers selection engine, as described in detail in examples 1-5 below:

"BONE" for "bone";

"COL" for "colon";

"EPI" for "epithelial";

"GEN" for "general";

"LIVER" for "liver";

"LUN" for "lung";

"LYMPH" for "lymph nodes";

"MARROW" for "bone marrow";

"OVA" for "ovary";

"PANCREAS" for "pancreas";

"PRO" for "prostate";

"STOMACH" for "stomach";

"TCELL" for "T cells";

"THYROID" for "Thyroid";

"MAM" for "breast";

"BRAIN" for "brain";

"UTERUS" for "uterus";

"SKIN" for "skin";

"KIDNEY" for "kidney";
"MUSCLE" for "muscle";
"ADREN" for "adrenal";
"HEAD" for "head and neck";
"BLADDER" for "bladder";

It should be noted that the terms "segment", "seg" and "node" are used interchangeably in reference to nucleic acid sequences of the present invention; they refer to portions of nucleic acid sequences that were shown to have one or more properties as described below. They are also the building blocks that were used to construct complete nucleic acid sequences as described in greater detail below. Optionally and preferably, they are examples of oligonucleotides which are embodiments of the present invention, for example as amplicons, hybridization units and/or from which primers and/or complementary oligonucleotides may optionally be derived, and/or for any other use.

As used herein the phrase "lung cancer" refers to cancers of the lung including small cell lung cancer and non-small cell lung cancer, including but not limited to lung adenocarcinoma, squamous cell carcinoma, and adenocarcinoma.

The term "marker" in the context of the present invention refers to a nucleic acid fragment, a peptide, or a polypeptide, which is differentially present in a sample taken from subjects (patients) having lung cancer (or one of the above indicative conditions) as compared to a comparable sample taken from subjects who do not have lung cancer (or one of the above indicative conditions).

The phrase "differentially present" refers to differences in the quantity of a marker present in a sample taken from patients having lung cancer (or one of the above indicative conditions) as compared to a comparable sample taken from patients who do not have lung cancer (or one of the above indicative conditions). For example, a nucleic acid fragment may optionally be differentially present between the two samples if the amount of the nucleic acid fragment in one sample is significantly different from the amount of the nucleic acid fragment in the other sample, for example as measured by hybridization and/or NAT-based assays. A polypeptide is differentially present between the two samples if the amount of the polypeptide in one sample is significantly different from the amount of the polypeptide in the other sample. It

should be noted that if the marker is detectable in one sample and not detectable in the other, then such a marker can be considered to be differentially present.

As used herein the phrase "diagnostic" means identifying the presence or nature of a pathologic condition. Diagnostic methods differ in their sensitivity and specificity. The "sensitivity" of a diagnostic assay is the percentage of diseased individuals who test positive (percent of "true positives"). Diseased individuals not detected by the assay are "false negatives." Subjects who are not diseased and who test negative in the assay are termed "true negatives." The "specificity" of a diagnostic assay is 1 minus the false positive rate, where the "false positive" rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

As used herein the phrase "diagnosing" refers to classifying a disease or a symptom, determining a severity of the disease, monitoring disease progression, forecasting an outcome of a disease and/or prospects of recovery. The term "detecting" may also optionally encompass any of the above.

Diagnosis of a disease according to the present invention can be effected by determining a level of a polynucleotide or a polypeptide of the present invention in a biological sample obtained from the subject, wherein the level determined can be correlated with predisposition to, or presence or absence of the disease. It should be noted that a "biological sample obtained from the subject" may also optionally comprise a sample that has not been physically removed from the subject, as described in greater detail below.

As used herein, the term "level" refers to expression levels of RNA and/or protein or to DNA copy number of a marker of the present invention.

Typically the level of the marker in a biological sample obtained from the subject is different (i.e., increased or decreased) from the level of the same variant in a similar sample obtained from a healthy individual (examples of biological samples are described herein).

Numerous well known tissue or fluid collection methods can be utilized to collect the biological sample from the subject in order to determine the level of DNA, RNA and/or polypeptide of the variant of interest in the subject.

Examples include, but are not limited to, fine needle biopsy, needle biopsy, core needle biopsy and surgical biopsy (e.g., brain biopsy), and lavage. Regardless of the procedure

employed, once a biopsy/sample is obtained the level of the variant can be determined and a diagnosis can thus be made.

Determining the level of the same variant in normal tissues of the same origin is preferably effected along-side to detect an elevated expression and/or amplification and/or a decreased expression, of the variant as opposed to the normal tissues.

A "test amount" of a marker refers to an amount of a marker in a subject's sample that is consistent with a diagnosis of lung cancer (or one of the above indicative conditions). A test amount can be either in absolute amount (e.g., microgram/ml) or a relative amount (e.g., relative intensity of signals).

A "control amount" of a marker can be any amount or a range of amounts to be compared against a test amount of a marker. For example, a control amount of a marker can be the amount of a marker in a patient with lung cancer (or one of the above indicative conditions) or a person without lung cancer (or one of the above indicative conditions). A control amount can be either in absolute amount (e.g., microgram/ml) or a relative amount (e.g., relative intensity of signals).

"Detect" refers to identifying the presence, absence or amount of the object to be detected.

A "label" includes any moiety or item detectable by spectroscopic, photo chemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, ^{35}S, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin-streptavadin, dioxigenin, haptens and proteins for which antisera or monoclonal antibodies are available, or nucleic acid molecules with a sequence complementary to a target. The label often generates a measurable signal, such as a radioactive, chromogenic, or fluorescent signal, that can be used to quantify the amount of bound label in a sample. The label can be incorporated in or attached to a primer or probe either covalently, or through ionic, van der Waals or hydrogen bonds, e.g., incorporation of radioactive nucleotides, or biotinylated nucleotides that are recognized by streptavadin. The label may be directly or indirectly detectable. Indirect detection can involve the binding of a second label to the first label, directly or indirectly. For example, the label can be the ligand of a binding partner, such as biotin, which is a binding partner for streptavadin, or a nucleotide sequence, which is the binding partner for a complementary sequence, to which it can specifically hybridize. The binding partner may itself

be directly detectable, for example, an antibody may be itself labeled with a fluorescent molecule. The binding partner also may be indirectly detectable, for example, a nucleic acid having a complementary nucleotide sequence can be a part of a branched DNA molecule that is in turn detectable through hybridization with other labeled nucleic acid molecules (see, e.g., P. D. Fahrlander and A. Klausner, Bio/Technology 6:1165 (1988)). Quantitation of the signal is achieved by, e.g., scintillation counting, densitometry, or flow cytometry.

Exemplary detectable labels, optionally and preferably for use with immunoassays, include but are not limited to magnetic beads, fluorescent dyes, radiolabels, enzymes (e.g., horse radish peroxide, alkaline phosphatase and others commonly used in an ELISA), and calorimetric labels such as colloidal gold or colored glass or plastic beads. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound marker-specific antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the marker are incubated simultaneously with the mixture.

"Immunoassay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide (or other epitope), refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times greater than the background (non-specific signal) and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to seminal basic protein from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with seminal basic protein and not with other proteins, except for polymorphic variants and alleles of seminal basic protein. This selection may be achieved by subtracting out antibodies that cross-react with seminal basic protein molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a

particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, Antibodies, A Laboratory Manual (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

According to preferred embodiments of the present invention, preferably any of the above nucleic acid and/or amino acid sequences further comprises any sequence having at least about 70%, preferably at least about 80%, more preferably at least about 90%, most preferably at least about 95% homology thereto.

Unless otherwise noted, all experimental data relates to variants of the present invention, named according to the segment being tested (as expression was tested through RT-PCR as described).

All nucleic acid sequences and/or amino acid sequences shown herein as embodiments of the present invention relate to their isolated form, as isolated polynucleotides (including for all transcripts), oligonucleotides (including for all segments, amplicons and primers), peptides (including for all tails, bridges, insertions or heads, optionally including other antibody epitopes as described herein) and/or polypeptides (including for all proteins). It should be noted that oligonucleotide and polynucleotide, or peptide and polypeptide, may optionally be used interchangeably.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1 and 2.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1022, 1023, 1024, 1025, 1026 and 1027.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1281 and 1282.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 3 and 4.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037 and 1038.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1283 and 1284.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 5, 6, 7 and 8.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065 and 1066.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1285, 1286, 1287 and 1288.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 9, 10, 11, 12, 13, 14 and 15.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099 and 1100.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1289, 1290, 1291, 1292, 1293 and 1294.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 20 and 21.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1130, 1131, 1132, 1133 and 1134.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1299 and 1300.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 22, 23 and 24.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143 and 1144.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1301, 1302 and 1303.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 25, 26 and 27.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155 and 1156.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1304 and 1305.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 28.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170 and 1171.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1306.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 29 and 30.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190 and 1191.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1307 and 1308.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 31.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1192, 1193, 1194, 1195, 1196, 1197 and 1198.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1309.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 32.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1199, 1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210, 1211, 1212, 1213, 1214 and 1215.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO. 1310.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 33.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1216 and 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1225, 1226 and 1227.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1311.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 34.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1228, 1229, 1230, 1231, 1232 and 1223.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1312.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 35.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243, 1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253 and 1254.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1313.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 36, 37, 38, 39 and 40.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265, 1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274 and 1275.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1314, 1315, 1316 and 1317.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 125, 126, 127, 128, 129 and 130.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901 and 902.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1394, 1395, 1396, 1397 and 1398.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising a transcript SEQ ID NOs: 131 and 132.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 903, 904, 905, 906, 907, 907, 908 and 909.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1399 and 1400.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 99, 100, 101 and 102.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787 and 788.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1372, 1373, 1374 and 1375.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 134.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935 and 936.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1402.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NO: 133.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 910, 911 and 912.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 141, 142 and 142.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989 and 990.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising :

Protein Name

HUMOSTRO_PEA_1_PEA_1_P21

HUMOSTRO_PEA_1_PEA_1_P25

HUMOSTRO_PEA_1_PEA_1_P30

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 51, 52, 53,, 54, 55, 56 and 57.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547,548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563,, 564, 565, 566, 567, 568, 569 and 570.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1327, 1328, 1329, 1330, 1331, 1332 and 1333.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 135, 136, 137, 138, 139 and 140.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959 and 960.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1403, 1404, 1405, 1406, 1407 and 1408.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 41, 42, 43, 44, 45, 46 and 47..

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 482, 483, 484, 495, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500 and 501.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1318, 1319, 1320, 1321, 1322 and 1323.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 121, 122, 123 and 124.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 876, 877, 878, 879, 880, 881, 882, 883, 884, 885 and 886.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1390, 1391, 1392 and 1393.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 48, 49 and 50.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516 and 517.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1324, 1325 and 1326.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1464 and 1465.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising a SEQ ID NOs: 1276, 1277, 1278, 1279 and 1280.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1415.

Protein Name	Corresponding Transcript(s)
HSU33147_PEA_1_P5	HSU33147_PEA_1_T1; HSU33147_PEA_1_T2

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NO: 58.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 571, 572, 573, 574, 575, 576, 577 and 578.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1334.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 74, 75, 76, 77, 78, 79, 80, 81 and 82.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692 and 693.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1350, 1351, 1352, 1353, 1354, 1355, 1356 and 1357.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs:

Transcript Name

T23580_T10

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 579, 580, 581, 582 and 583.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1335.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 59, 60, 61, 62, 63 and 64.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614 and 615.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1336, 1337, 1338, 1339 and 1340.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 65, 66, 67, 68, 69, 70, 71, 72 and 73.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658 and 659.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1341, 1342, 1343, 1344, 1345, 1346, 1347, 1348 and 1349.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95 and 96.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 695, 696, 697, 698, 699, 700, 701, 702, 703, 704 and 705.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1367, 1368 and 1369.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 97 and 98.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740 and 741.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1370 and 1371.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 103, 104, 105, 106, 107 and 108.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 8 09, 810, 811, 812 and 813.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1376, 1377, 1378 and 1379.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 114, 115, 116, 117, 118 and 119.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874 and 875.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1385, 1386, 1387, 1388 and 1389.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 144, 145, 146, 147, 148 and 149.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015 and 1016.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs: 1409, 1410, 1411, 1412 and 1413.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NO: 150.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 1017, 1018, 1019, 1020 and 1021.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NO: 1414.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 109, 110, 111, 112 and 113.

According to preferred embodiments of the present invention, there is provided an isolated polynucleotide comprising SEQ ID NOs: 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854 and 855.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide comprising SEQ ID NOs 1380, 1381, 1382, 1383 and 1384.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for.HSSTROL3_P4, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P4, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P4, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRFPVHAALVWGPE KNKIYFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADG corresponding to amino acids 165 - 445 of MM11_HUMAN, which also corresponds to amino acids 165 - 445 of HSSTROL3_P4, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ALGVRQLVGGGHSSRFSHLVVAGLPHACHRKSGSSSQVLCPEPSALLSVAG corresponding to amino acids 446 - 496 of HSSTROL3_P4, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HSSTROL3_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ALGVRQLVGGGHSSRFSHLVVAGLPHACHRKSGSSSQVLCPEPSALLSVAG in HSSTROL3_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSSTROL3_P5, comprising a first amino acid sequence being at least 90 % homologous to

MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P5, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P5, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQ corresponding to amino acids 165 - 358 of MM11_HUMAN, which also corresponds to amino acids 165 - 358 of HSSTROL3_P5, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ELGFPSSTGRDESLEHCRCQGLHK corresponding to amino acids 359 - 382 of HSSTROL3_P5, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HSSTROL3_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ELGFPSSTGRDESLEHCRCQGLHK in HSSTROL3_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSSTROL3_P7, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P7, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P7, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG

LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQG corresponding to amino acids 165 - 359 of MM11_HUMAN, which also corresponds to amino acids 165 - 359 of HSSTROL3_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TTGVSTPAPGV corresponding to amino acids 360 - 370 of HSSTROL3_P7, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HSSTROL3_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TTGVSTPAPGV in HSSTROL3_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSSTROL3_P8, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P8, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P8, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLE corresponding to amino acids 165 - 286 of MM11_HUMAN, which also corresponds to amino acids 165 - 286 of HSSTROL3_P8, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRPCLPVPLLLCWPL corresponding to amino acids 287 - 301 of HSSTROL3_P8, wherein

said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HSSTROL3_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRPCLPVPLLLCWPL in HSSTROL3_P8.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSSTROL3_P9, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQK corresponding to amino acids 1 - 96 of MM11_HUMAN, which also corresponds to amino acids 1 - 96 of HSSTROL3_P9, a second amino acid sequence being at least 90 % homologous to RILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 113 - 163 of MM11_HUMAN, which also corresponds to amino acids 97 - 147 of HSSTROL3_P9, a bridging amino acid H corresponding to amino acid 148 of HSSTROL3_P9, a third amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQG corresponding to amino acids 165 - 359 of MM11_HUMAN, which also corresponds to amino acids 149 - 343 of HSSTROL3_P9, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TTGVSTPAPGV corresponding to amino acids 344 - 354 of HSSTROL3_P9, wherein said first amino acid sequence, second amino acid sequence, bridging amino acid, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HSSTROL3_P9, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally

at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KR, having a structure as follows: a sequence starting from any of amino acid numbers 96-x to 96; and ending at any of amino acid numbers 97+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HSSTROL3_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TTGVSTPAPGV in HSSTROL3_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCA1XIA_P14, comprising a first amino acid sequence being at least 90 % homologous to

MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT
GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY
NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM
IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH
YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT
EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED
TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN
GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG
RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM
GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP
GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAG
PRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQG
PIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPPGPQGPIGYPGPRGVK
GADGVRGLKGSKGEKGEDGFPGFKGDMGLKGDRGEVGQIGPRGEDGPEGPKGRAGPT
GDPGPSGQAGEKGKLGVPGLPGYPGRQGPKGSTGFPGFPGANGEKGARGVAGKPGPR
GQRGPTGPRGSRGARGPTGKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGP
PGKDGLPGHPGQRGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQG
LPGAAGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQGPPGP

V corresponding to amino acids 1 - 1056 of CA1B_HUMAN_V5, which also corresponds to amino acids 1 - 1056 of HUMCA1XIA_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSMMIINSQTIMVVNYSSSFITLML corresponding to amino acids 1057 - 1081 of HUMCA1XIA_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCA1XIA_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSMMIINSQTIMVVNYSSSFITLML in HUMCA1XIA_P14.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCA1XIA_P15, comprising a first amino acid sequence being at least 90 % homologous to MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAG PRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQG PIGPPGEK corresponding to amino acids 1 - 714 of CA1B_HUMAN, which also corresponds to amino acids 1 - 714 of HUMCA1XIA_P15, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

MCCNLSFGILIPLQK corresponding to amino acids 715 - 729 of HUMCA1XIA_P15, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCA1XIA_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MCCNLSFGILIPLQK in HUMCA1XIA_P15.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCA1XIA_P16, comprising a first amino acid sequence being at least 90 % homologous to
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT
GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY
NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM
IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH
YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT
EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED
TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN
GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG
RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM
GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP
GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEA
corresponding to amino acids 1 - 648 of CA1B_HUMAN, which also corresponds to amino acids 1 - 648 of HUMCA1XIA_P16, a second amino acid sequence being at least 90 % homologous to GMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQGPIGPPGEK corresponding to amino acids 667 - 714 of CA1B_HUMAN, which also corresponds to amino acids 649 - 696 of HUMCA1XIA_P16, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSFSFSLFYKKVIKFACDKRFVGRHDERKVVKLSLPLYLIYE corresponding to amino

acids 697 - 738 of HUMCA1XIA_P16, wherein said first amino acid sequence, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HUMCA1XIA_P16, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise AG, having a structure as follows: a sequence starting from any of amino acid numbers 648-x to 648; and ending at any of amino acid numbers 649+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCA1XIA_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSFSFSLFYKKVIKFACDKRFVGRHDERKVVKLSLPLYLIYE in HUMCA1XIA_P16.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCA1XIA_P17, comprising a first amino acid sequence being at least 90 % homologous to MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH YSPDCDSSAPKAAQAQEPQIDE corresponding to amino acids 1 - 260 of CA1B_HUMAN, which also corresponds to amino acids 1 - 260 of HUMCA1XIA_P17, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRSTRPEKVFVFQ corresponding to amino acids 261 - 273 of HUMCA1XIA_P17, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCA1XIA_P17, comprising a polypeptide being

at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRSTRPEKVFVFQ in HUMCA1XIA_P17.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R20779_P2, comprising a first amino acid sequence being at least 90 % homologous to MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNTAEIQHCLV NAGDVGCGVFECFENNSCEIRGLHGICMTFLHNAGKFDAQGKSFIKDALKCKAHALRH RFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQENTRVIVEMIHFKDLLLHE corresponding to amino acids 1 - 169 of STC2_HUMAN, which also corresponds to amino acids 1 - 169 of R20779_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence CYKIEITMPKRRKVKLRD corresponding to amino acids 170 - 187 of R20779_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R20779_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence CYKIEITMPKRRKVKLRD in R20779_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P21, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQKQNLLAPQ corresponding to amino acids 1 - 58 of OSTP_HUMAN, which also corresponds to amino acids 1 - 58 of HUMOSTRO_PEA_1_PEA_1_P21, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VFLNFS corresponding to amino acids 59 - 64 of HUMOSTRO_PEA_1_PEA_1_P21, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMOSTRO_PEA_1_PEA_1_P21, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VFLNFS in HUMOSTRO_PEA_1_PEA_1_P21.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P25, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQ corresponding to amino acids 1 - 31 of OSTP_HUMAN, which also corresponds to amino acids 1 - 31 of HUMOSTRO_PEA_1_PEA_1_P25, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence H corresponding to amino acids 32 - 32 of HUMOSTRO_PEA_1_PEA_1_P25, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P30, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQ corresponding to amino acids 1 - 31 of OSTP_HUMAN, which also corresponds to amino acids 1 - 31 of HUMOSTRO_PEA_1_PEA_1_P30, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSIFYVFI corresponding to amino acids 32 - 39 of HUMOSTRO_PEA_1_PEA_1_P30, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMOSTRO_PEA_1_PEA_1_P30, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSIFYVFI in HUMOSTRO_PEA_1_PEA_1_P30.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P10, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISE corresponding to amino acids 1 - 67 of PLTP_HUMAN, which also corresponds to amino acids 1 - 67 of HUMPHOSLIP_PEA_2_P10, and a second amino acid sequence being at least 90 % homologous to KVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMK DPVASTSNLDMDFRGAFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMES YFRAGALQLLLVGDKVPHDLDMLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKP SGTTISVTASVTIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSN HSALESLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVTNHAGFLTI GADLHFAKGLREVIEKNRPADVRASTAPTPSTAAV corresponding to amino acids 163 - 493 of PLTP_HUMAN, which also corresponds to amino acids 68 - 398 of HUMPHOSLIP_PEA_2_P10, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HUMPHOSLIP_PEA_2_P10, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EK, having a structure as follows: a sequence starting from any of amino acid numbers 67-x to 67; and ending at any of amino acid numbers 68+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P12, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRG

AFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDK VPHDLDMLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASVTIALVP PDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHSALESLALIPLQAPLK TMLQIGVMPMLN corresponding to amino acids 1 - 427 of PLTP_HUMAN, which also corresponds to amino acids 1 - 427 of HUMPHOSLIP_PEA_2_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKAGV corresponding to amino acids 428 - 432 of HUMPHOSLIP_PEA_2_P12, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKAGV in HUMPHOSLIP_PEA_2_P12.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P31, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISE corresponding to amino acids 1 - 67 of PLTP_HUMAN, which also corresponds to amino acids 1 - 67 of HUMPHOSLIP_PEA_2_P31, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PGLERGADKFPVVGGSSLFLALDLTLRPPVG corresponding to amino acids 68 - 98 of HUMPHOSLIP_PEA_2_P31, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P31, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PGLERGADKFPVVGGSSLFLALDLTLRPPVG in HUMPHOSLIP_PEA_2_P31.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P33, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQ corresponding to amino acids 1 - 183 of PLTP_HUMAN, which also corresponds to amino acids 1 - 183 of HUMPHOSLIP_PEA_2_P33, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VWAATGRRVARVGMLSL corresponding to amino acids 184 - 200 of HUMPHOSLIP_PEA_2_P33, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P33, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VWAATGRRVARVGMLSL in HUMPHOSLIP_PEA_2_P33.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P34, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQICPVLYHAGTVLLNSLLDTVPV corresponding to amino acids 1 - 205 of PLTP_HUMAN, which also corresponds to amino acids 1 - 205 of HUMPHOSLIP_PEA_2_P34, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LWTSLLALTIPS corresponding to amino acids 206 - 217 of HUMPHOSLIP_PEA_2_P34, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P34, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LWTSLLALTIPS in HUMPHOSLIP_PEA_2_P34.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P35, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWF corresponding to amino acids 1 - 109 of PLTP_HUMAN, which also corresponds to amino acids 1 - 109 of HUMPHOSLIP_PEA_2_P35, a second amino acid sequence bridging amino acid sequence comprising of L, a third amino acid sequence being at least 90 % homologous to KVYDFLSTFITSGMRFLLNQQ corresponding to amino acids 163 - 183 of PLTP_HUMAN, which also corresponds to amino acids 111 - 131 of HUMPHOSLIP_PEA_2_P35, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VWAATGRRVARVGMLSL corresponding to amino acids 132 - 148 of HUMPHOSLIP_PEA_2_P35, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for an edge portion of HUMPHOSLIP_PEA_2_P35, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise FLK having a structure as follows (numbering according to HUMPHOSLIP_PEA_2_P35): a sequence starting from any of amino acid numbers 109-x to 109; and ending at any of amino acid numbers 111 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P35, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VWAATGRRVARVGMLSL in HUMPHOSLIP_PEA_2_P35.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P6, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGG LPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESIEKISKVEC GFAT corresponding to amino acids 1 - 412 of CT31_HUMAN, which also corresponds to amino acids 1 - 412 of R38144_PEA_2_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LASFSHMSDQRSARPQAGQPHGVVLPGRDCEIPLPPV corresponding to amino acids 413 - 449 of R38144_PEA_2_P6, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LASFSHMSDQRSARPQAGQPHGVVLPGRDCEIPLPPV in R38144_PEA_2_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P13, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD

ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQ corresponding to amino acids 1 - 323 of CT31_HUMAN, which also corresponds to amino acids 1 - 323 of R38144_PEA_2_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence NLLKAQCTSTVPRGIPPS corresponding to amino acids 324 - 341 of R38144_PEA_2_P13, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence NLLKAQCTSTVPRGIPPS in R38144_PEA_2_P13.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P15, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLE corresponding to amino acids 1 - 282 of CT31_HUMAN, which also corresponds to amino acids 1 - 282 of R38144_PEA_2_P15, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PHWRH corresponding to amino acids 283 - 287 of R38144_PEA_2_P15, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PHWRH in R38144_PEA_2_P15.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P19, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGG LPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESIEKISKVEC GFAT corresponding to amino acids 1 - 412 of CT31_HUMAN, which also corresponds to amino acids 1 - 412 of R38144_PEA_2_P19, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence KRSRSVAQAGVQWCDHDSPQP corresponding to amino acids 413 - 433 of R38144_PEA_2_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P19, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence KRSRSVAQAGVQWCDHDSPQP in R38144_PEA_2_P19.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P24, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD

ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIR corresponding to amino acids 1 - 121 of CT31_HUMAN, which also corresponds to amino acids 1 - 121 of R38144_PEA_2_P24, and a second amino acid sequence being at least 90 % homologous to

EYNKAIRNYTRFDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLN YYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDA VESIEKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNNGSTFDA VITPYGECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVEDLMREFYSLKRSRSKFQ KNTVSSGPWEPPARPGTLFSPENHDQARERKPAKQKVPLLSCPSQPFTSKLALLGQVFL DSS corresponding to amino acids 282 - 578 of CT31_HUMAN, which also corresponds to amino acids 122 - 418 of R38144_PEA_2_P24, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of R38144_PEA_2_P24, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise RE, having a structure as follows: a sequence starting from any of amino acid numbers 121-x to 121; and ending at any of amino acid numbers 122+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to

MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR corresponding to amino acids 1 - 36 of AAH16184, which also corresponds to amino acids 1 - 36 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence FWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 37 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence FWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHY corresponding to amino acids 1 - 35 of AAQ88943, which also corresponds to amino acids 1 - 35 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RFWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 36 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RFWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR corresponding to amino acids 1 - 36 of CT31_HUMAN, which also corresponds to amino acids 1 - 36 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence FWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 37 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence FWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for AA161187_P6, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR corresponding to amino acids 1 - 42 of AA161187_P6, and a second amino acid sequence being at least 90 % homologous to
GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYS
DLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPV
TYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNH
LFLKYSFRKDIFGDMVCAGNAQGGKDACFGDSGGPLACNKNGLWYQIGVVSWGVGC
GRPNRPGVYTNISHHFEWIQKLMAQSGMSQPDPSWPLLFFPLLWALPLLGPV
corresponding to amino acids 31 - 314 of TEST_HUMAN, which also corresponds to amino acids 43 - 326 of AA161187_P6, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of AA161187_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR of AA161187_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for AA161187_P13, comprising a first amino acid sequence being at least 90 % homologous to
MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS
LRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY
YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG

WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GSSGRHHKQLYVQPPLPQVQFPQGHLWRHG corresponding to amino acids 184 - 213 of AA161187_P13, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of AA161187_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GSSGRHHKQLYVQPPLPQVQFPQGHLWRHG in AA161187_P13.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for AA161187_P14, comprising a first amino acid sequence being at least 90 % homologous to MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS LRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GCCLSPSHYRPHSTAISPHPPGSSGRHHKQLYVQPPLPQVQFPQGHLWRHGLCWQCPRR EGCLLRECPCHHSQPRKASCVPVPYLTLMPTPGGGDCCPTLQMQKRRLGCCQGEEEDV HPVYPAP corresponding to amino acids 184 - 307 of AA161187_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of AA161187_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GCCLSPSHYRPHSTAISPHPPGSSGRHHKQLYVQPPLPQVQFPQGHLWRHGLCWQCPRR

EGCLLRECPCHHSQPRKASCVPVP YLTLMPTPGGGDCCPTLQMQKRRLGCCQGEEEDV HPVYPAP in AA161187_P14.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for AA161187_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR corresponding to amino acids 1 - 42 of AA161187_P18, a second amino acid sequence being at least 90 % homologous to GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFET corresponding to amino acids 31 - 86 of TEST_HUMAN, which also corresponds to amino acids 43 - 98 of AA161187_P18, a third amino acid sequence being at least 90 % homologous to DLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPV TYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNH LFLKYSFRKDIFGDMVCAGNAQGGKDACF corresponding to amino acids 89 - 235 of TEST_HUMAN, which also corresponds to amino acids 99 - 245 of AA161187_P18, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSVPATTPSPGKHPVSLCLI corresponding to amino acids 246 - 265 of AA161187_P18, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of AA161187_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR of AA161187_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of AA161187_P18, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TD, having a structure as follows: a

sequence starting from any of amino acid numbers 98-x to 99; and ending at any of amino acid numbers 99+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of AA161187_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSVPATTPSPGKHPVSLCLI in AA161187_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for AA161187_P19, comprising a first amino acid sequence being at least 90 % homologous to MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS LRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P19, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DKRTQ corresponding to amino acids 184 - 188 of AA161187_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of AA161187_P19, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DKRTQ in AA161187_P19.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z25299_PEA_2_P2, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLK CCMGMCGKSCVSPVK corresponding to amino acids 1 - 131 of ALK1_HUMAN, which also corresponds to amino acids 1 - 131 of Z25299_PEA_2_P2, and a second amino acid sequence

being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKQGMRAH corresponding to amino acids 132 - 139 of Z25299_PEA_2_P2, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z25299_PEA_2_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKQGMRAH in Z25299_PEA_2_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z25299_PEA_2_P3, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLK CCMGMCGKSCVSPVK corresponding to amino acids 1 - 131 of ALK1_HUMAN, which also corresponds to amino acids 1 - 131 of Z25299_PEA_2_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GEKRHHKQLRDQEVDPLEMRRHSAG corresponding to amino acids 132 - 156 of Z25299_PEA_2_P3, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z25299_PEA_2_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GEKRHHKQLRDQEVDPLEMRRHSAG in Z25299_PEA_2_P3.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z25299_PEA_2_P7, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNP corresponding to amino acids 1 - 81 of ALK1_HUMAN,

which also corresponds to amino acids 1 - 81 of Z25299_PEA_2_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGSLGSAQ corresponding to amino acids 82 - 89 of Z25299_PEA_2_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z25299_PEA_2_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGSLGSAQ in Z25299_PEA_2_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z25299_PEA_2_P10, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPT corresponding to amino acids 1 - 82 of ALK1_HUMAN, which also corresponds to amino acids 1 - 82 of Z25299_PEA_2_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R66178_P3, comprising a first amino acid sequence being at least 90 % homologous to MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLEL EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT corresponding to amino acids 1 - 334 of PVR1_HUMAN, which also corresponds to amino acids 1 - 334 of R66178_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GEGHSLPISPGVLQTQNCGP corresponding to amino acids 335 - 354 of R66178_P3, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R66178_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GEGHSLPISPGVLQTQNCGP in R66178_P3.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R66178_P4, comprising a first amino acid sequence being at least 90 % homologous to MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLEL EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT corresponding to amino acids 1 - 334 of PVR1_HUMAN, which also corresponds to amino acids 1 - 334 of R66178_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence AFCQLIYPGKGRTRARMF corresponding to amino acids 335 - 352 of R66178_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R66178_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence AFCQLIYPGKGRTRARMF in R66178_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R66178_P8, comprising a first amino acid sequence being at least 90 % homologous to MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVS VLAPYRERVEFLRPSFTDGTIRLSRLEL EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS

ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVE corresponding to amino acids 1 - 330 of PVR1_HUMAN, which also corresponds to amino acids 1 - 330 of R66178_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence NSPTPRLLPNMGGAPGRCPRPSLGAWRGASCWC corresponding to amino acids 331 - 363 of R66178_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R66178_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence NSPTPRLLPNMGGAPGRCPRPSLGAWRGASCWC in R66178_P8.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSU33147_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAI DELKECFLNQTDETLSNVE corresponding to amino acids 1 - 78 of MGBA_HUMAN, which also corresponds to amino acids 1 - 78 of HSU33147_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to QLIYDSSLCDLF corresponding to amino acids 82 - 93 of MGBA_HUMAN, which also corresponds to amino acids 79 - 90 of HSU33147_PEA_1_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HSU33147_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EQ, having a

structure as follows: a sequence starting from any of amino acid numbers 78-x to 78; and ending at any of amino acid numbers 79+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSU33147_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAI DELKECFLNQTDETLSNVE corresponding to amino acids 1 - 78 of MGBA_HUMAN, which also corresponds to amino acids 1 - 78 of HSU33147_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to QLIYDSSLCDLF corresponding to amino acids 82 - 93 of MGBA_HUMAN, which also corresponds to amino acids 79 - 90 of HSU33147_PEA_1_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HSU33147_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EQ, having a structure as follows: a sequence starting from any of amino acid numbers 78-x to 78; and ending at any of amino acid numbers 79+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P3, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD

QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKD corresponding to amino acids 1 - 517 of APP1_HUMAN, which also corresponds to amino acids 1 - 517 of M78076_PEA_1_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GE corresponding to amino acids 518 - 519 of M78076_PEA_1_P3, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to

MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKG

corresponding to amino acids 1 - 526 of APP1_HUMAN, which also corresponds to amino acids 1 - 526 of M78076_PEA_1_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ECLTVNPSLQIPLNP corresponding to amino acids 527 - 541 of M78076_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ECLTVNPSLQIPLNP in M78076_PEA_1_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P12, comprising a first amino acid sequence being at least 90 % homologous to

MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKG

corresponding to amino acids 1 - 526 of APP1_HUMAN, which also corresponds to amino acids 1 - 526 of M78076_PEA_1_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ECVCSKGFPFPLIGDSEG corresponding to amino acids 527 - 544 of M78076_PEA_1_P12, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ECVCSKGFPFPLIGDSEG in M78076_PEA_1_P12.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to

MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG

SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGST
EQDAASPEKEKMNPLEQYERKVNASVPRGFPFHSSEIQRDEL corresponding to amino acids 1 - 570 of APP1_HUMAN, which also corresponds to amino acids 1 - 570 of M78076_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRGGTAGYLGEETRGQRPGCDSQSHTGPSKKPSAPSPLPAGTSWDRGVP corresponding to amino acids 571 - 619 of M78076_PEA_1_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRGGTAGYLGEETRGQRPGCDSQSHTGPSKKPSAPSPLPAGTSWDRGVP in M78076_PEA_1_P14.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P21, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN E corresponding to amino acids 1 - 352 of APP1_HUMAN, which also corresponds to amino acids 1 - 352 of M78076_PEA_1_P21, and a second amino acid sequence being at least 90 % homologous to

AERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMT
LPKGSTEQDAASPEKEKMNPLEQYERKVNASVPRGFPFHSSEIQRDELAPAGTGVSREA
VSGLLIMGAGGGSLIVLSMLLLRRKKPYGAISHGVVEVDPMLTLEEQQLRELQRHGYE
NPTYRFLEERP corresponding to amino acids 406 - 650 of APP1_HUMAN, which also corresponds to amino acids 353 - 597 of M78076_PEA_1_P21, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of M78076_PEA_1_P21, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EA, having a structure as follows: a sequence starting from any of amino acid numbers 352-x to 352; and ending at any of amino acid numbers 353+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P24, comprising a first amino acid sequence being at least 90 % homologous to
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQI corresponding to amino acids 1 - 481 of APP1_HUMAN, which also corresponds to amino acids 1 - 481 of M78076_PEA_1_P24, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RECLLPWLPLQISEGRS corresponding to amino acids 482 - 498 of M78076_PEA_1_P24,

wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P24, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RECLLPWLPLQISEGRS in M78076_PEA_1_P24.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQV corresponding to amino acids 1 - 449 of APP1_HUMAN, which also corresponds to amino acids 1 - 449 of M78076_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LTSFQLPNAPLFLRRPRLRLFSCPLDPLSVSWTPSYPLNTASLPLPSLSAQLPDPETWTLT CCVFDPCFLALGFLLPPPSILCSVPWIFTAFPRIVFFFFFFLRQVLALSPRQESSVRSWLIAT STSWVQAILLPQPLE corresponding to amino acids 450 - 588 of M78076_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

LTSFQLPNAPLFLRRPRLRLFSCPLDPLSVSWTPSYPLNTASLPLPSLSAQLPDPETWTLT
CCVFDPCFLALGFLLPPPSILCSVPWIFTAFPRIVFFFFFFLRQVLALSPRQESSVRSWLIAT
STSWVQAILLPQPLE in M78076_PEA_1_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M78076_PEA_1_P25, comprising a first amino acid sequence being at least 90 % homologous to
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQ corresponding to amino acids 1 - 448 of APP1_HUMAN, which also corresponds to amino acids 1 - 448 of M78076_PEA_1_P25, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence
PQNPNSQPRAAGSLEVIISHPFVRRLEILISPFQFQNSIPKNSQIVPAASPRGTSSP corresponding to amino acids 449 - 505 of M78076_PEA_1_P25, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M78076_PEA_1_P25, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PQNPNSQPRAAGSLEVIISHPFVRRLEILISPFQFQNSIPKNSQIVPAASPRGTSSP in M78076_PEA_1_P25.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M79217_PEA_1_P1, comprising a first amino acid sequence being at least 90 % homologous to
MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD

EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK
KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC
RLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYYVTENADIA
CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG
RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL
RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW
ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY
QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV
LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL
RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF
QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL
WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVVWREARD
RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRD
MVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFHERHKCINFF
VKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 13 - 931 of BAA25445, which also corresponds to amino acids 1 - 919 of M79217_PEA_1_P1.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M79217_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to

MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD
EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK
KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC
RLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYYVTENADIA
CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG
RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL
RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW
ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY
QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV
LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL
RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF
QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL

WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARD RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK corresponding to amino acids 1 - 807 of EXL3_HUMAN, which also corresponds to amino acids 1 - 807 of M79217_PEA_1_P2, and a second amino acid sequence being at least 90 % homologous to AIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFHERHK CINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 820 - 919 of EXL3_HUMAN, which also corresponds to amino acids 808 - 907 of M79217_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of M79217_PEA_1_P2, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KA, having a structure as follows: a sequence starting from any of amino acid numbers 807-x to 807; and ending at any of amino acid numbers 808+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M79217_PEA_1_P4, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PELRQPARLGLPECWDYRHEPRCPAQMGSHFIVQAGLKLLASSKPPKCWDY corresponding to amino acids 1 - 51 of M79217_PEA_1_P4, and a second amino acid sequence being at least 90 % homologous to RVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSY VMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFH ERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 759 - 919 of EXL3_HUMAN, which also corresponds to amino acids 52 - 212 of M79217_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of M79217_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PELRQPARLGLPECWDYRHEPRCPAQMGSHFIVQAGLKLLASSKPPKCWDY of M79217_PEA_1_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M79217_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC RLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYYVTENADIA CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARD RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK corresponding to amino acids 1 - 807 of EXL3_HUMAN, which also corresponds to amino acids 1 - 807 of M79217_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRKSW corresponding to amino acids 808 - 812 of M79217_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M79217_PEA_1_P8, comprising a polypeptide being

at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRKSW in M79217_PEA_1_P8.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P4, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MATYIH corresponding to amino acids 1 - 6 of M62096_PEA_1_P4, and a second amino acid sequence being at least 90 % homologous to VSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNC RTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKT LKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKE KYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQ IENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRE LSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYIS KMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQN MEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQ MESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREM KLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDD GGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALES ALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTA VHAIRGGGGSSSNSTHYQK corresponding to amino acids 239 - 957 of KF5C_HUMAN, which also corresponds to amino acids 7 - 725 of M62096_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of M62096_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MATYIH of M62096_PEA_1_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P5, comprising a first amino acid

sequence being at least 90 % homologous to
MTRILQDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWK
KKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNI
APVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRR
DYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDEL
AQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNG
VIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHE
AKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQ
DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDY
NKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTT
RVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRL
RATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPI
RPGHYPASSPTAVHAIRGGGGSSSNSTHYQK corresponding to amino acids 284 - 957 of KF5C_HUMAN, which also corresponds to amino acids 1 - 674 of M62096_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P3, comprising a first amino acid sequence being at least 90 % homologous to
MELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSL
YRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKD
EVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQELS
NHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKS
LVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQL
EESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAH
QKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLN
DKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQK
QKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKEN
AMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGG
SSSNSTHYQK corresponding to amino acids 365 - 957 of KF5C_HUMAN, which also corresponds to amino acids 1 - 593 of M62096_PEA_1_P3.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P7, comprising a first amino acid

sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MTQNFRLMWNILLFPLNFS corresponding to amino acids 1 - 19 of M62096_PEA_1_P7, and a second amino acid sequence being at least 90 % homologous to LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL QTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVR DNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRA KNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK corresponding to amino acids 738 - 957 of KF5C_HUMAN, which also corresponds to amino acids 20 - 239 of M62096_PEA_1_P7, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of M62096_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MTQNFRLMWNILLFPLNFS of M62096_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK TLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRL QIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYI SKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQN

MEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQ MESHREAHQKQLSRLRDEIEEKQKIIDEIR corresponding to amino acids 1 - 736 of KF5C_HUMAN, which also corresponds to amino acids 1 - 736 of M62096_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence E corresponding to amino acids 737 - 737 of M62096_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P9, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK TLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDE corresponding to amino acids 1 - 454 of KF5C_HUMAN, which also corresponds to amino acids 1 - 454 of M62096_PEA_1_P9, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VKNAIYFFFHKVLLLLFVVDVCSRNLIGIEAFHNYRIMWKFLGRCPFTASYKLIITEFRK corresponding to amino acids 455 - 514 of M62096_PEA_1_P9, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M62096_PEA_1_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

VKNAIYFFFHKVLLLLFVVDVCSRNLIGIEAFHNYRIMWKFLGRCPFTASYKLIITEFRK in M62096_PEA_1_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P10, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MTQNFRLMWNILLFPLNFS corresponding to amino acids 1 - 19 of M62096_PEA_1_P10, a second amino acid sequence being at least 90 % homologous to LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL QTLHNLRKLFVQDLTTRVKK corresponding to amino acids 738 - 815 of KF5C_HUMAN, which also corresponds to amino acids 20 - 97 of M62096_PEA_1_P10, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSSLCLNGTEKKIKDGREESFSVEISLA corresponding to amino acids 98 - 125 of M62096_PEA_1_P10, wherein said first amino acid sequence, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of M62096_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MTQNFRLMWNILLFPLNFS of M62096_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M62096_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSSLCLNGTEKKIKDGREESFSVEISLA in M62096_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD

HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE
EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE
KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN
CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK
TLKNVIQHLEMELNRWRN corresponding to amino acids 1 - 372 of KF5C_HUMAN, which also corresponds to amino acids 1 - 372 of M62096_PEA_1_P11, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DFLAAHVFGKLLE corresponding to amino acids 373 - 385 of M62096_PEA_1_P11, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M62096_PEA_1_P11, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DFLAAHVFGKLLE in M62096_PEA_1_P11.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M62096_PEA_1_P12, comprising a first amino acid sequence being at least 90 % homologous to
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ
EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD
HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE
EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE
KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN
CRTTIVICCSPSVFNEAETKSTLMFGQR corresponding to amino acids 1 - 323 of KF5C_HUMAN, which also corresponds to amino acids 1 - 323 of M62096_PEA_1_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence V corresponding to amino acids 324 - 324 of M62096_PEA_1_P12, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T99080_PEA_4_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MPASARLAGAGLLLAFLRALGCAGRAPGLS corresponding to amino acids 1 - 30 of T99080_PEA_4_P5, and a second amino acid sequence being at least 90 % homologous to MAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQGQLQGPIS KVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK corresponding to amino acids 1 - 99 of ACYO_HUMAN_V1, which also corresponds to amino acids 31 - 129 of T99080_PEA_4_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T99080_PEA_4_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MPASARLAGAGLLLAFLRALGCAGRAPGLS of T99080_PEA_4_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T99080_PEA_4_P8, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence M corresponding to amino acids 1 - 1 of T99080_PEA_4_P8, and a second amino acid sequence being at least 90 % homologous to QAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHIDKANFNNE KVILKLDYSDFQIVK corresponding to amino acids 28 - 99 of ACYO_HUMAN_V1, which also corresponds to amino acids 2 - 73 of T99080_PEA_4_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 90 % homologous to MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY

DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC GPVLTWME corresponding to amino acids 1 - 185 of SNXQ_HUMAN, which also corresponds to amino acids 1 - 185 of T08446_PEA_1_P18, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSW WRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVV DGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLY GKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNL AIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPA GRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEK QRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLS SQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGL GALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAP PAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTP ALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLR PGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEP LGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQ LRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQP SSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGM LGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPA SSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC corresponding to amino acids 186 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSW

WRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA
GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVV
DGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLY
GKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNL
AIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPA
GRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEK
QRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLS
SQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGL
GALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAP
PAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTP
ALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLR
PGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEP
LGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQ
LRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQP
SSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGM
LGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPA
SSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY
GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC in
T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM
PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRV corresponding to amino acids 1 - 443 of

T08446_PEA_1_P18, a second amino acid sequence being at least 90 % homologous to HDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVG MGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGSCPSTR LLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALGRG PSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHS SSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDG DELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRG PTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHLIPLLLRGA EAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALA LAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAW VPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSP CSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDR GENLYYEIGASEGSPYSG corresponding to amino acids 1 - 674 of Q9NT23, which also corresponds to amino acids 444 - 1117 of T08446_PEA_1_P18, a bridging amino acid P corresponding to amino acid 1118 of T08446_PEA_1_P18, and a third amino acid sequence being at least 90 % homologous to TRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPAR RPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHR VPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHS EGQTRSYC corresponding to amino acids 676 - 862 of Q9NT23, which also corresponds to amino acids 1119 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence, second amino acid sequence, bridging amino acid and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM

PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRV of T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM
PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANT
SMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTF
TSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAER
RKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRS
AKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSS
SESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDP
APPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGA
PASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLP
PPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS
LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRG
corresponding to amino acids 1 - 1010 of T08446_PEA_1_P18, and a second amino acid sequence being at least 90 % homologous to
LRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPG
LYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPP
DRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNL
ALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPL

LLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC corresponding to amino acids 1 - 295 of Q96CP3, which also corresponds to amino acids 1011 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP LLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANT SMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTF TSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAER RKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRS AKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSS SESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDP APPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGA PASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLP PPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRG of T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG

PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ corresponding to amino acids 1 - 154
of T08446_PEA_1_P18, a second amino acid sequence being at least 90 % homologous to
MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVI
KRYTAQAPDELSFEVGDIVSVIDMPPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPG
LKADADGPPCGIPAPQGISSLTSAVPRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVF
GCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPEL
SGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLP
PPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGAAAFR
EVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQ
ARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKP
LPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVG
PAPAGSCESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPR
CLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAA
LDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQ
QEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVA
EQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPR
QQSDGSLLRSQRPMGTSRRGLRGPA corresponding to amino acids 1 - 861 of BAC86902,
which also corresponds to amino acids 155 - 1015 of T08446_PEA_1_P18, a third amino acid
sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at
least 90% and most preferably at least 95% homologous to a polypeptide having the sequence
QVSAQLRAGGGGRDAPEAAAQSPCSVPS corresponding to amino acids 1016 - 1043 of
T08446_PEA_1_P18, a fourth amino acid sequence being at least 90 % homologous to
QVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLY
YEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPP
DHLGYS corresponding to amino acids 862 - 989 of BAC86902, which also corresponds to
amino acids 1044 - 1171 of T08446_PEA_1_P18, and a fifth amino acid sequence being at least
70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most
preferably at least 95% homologous to a polypeptide having the sequence
APQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAP
WGPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYP

TPSWSLHSEGQTRSYC corresponding to amino acids 1172 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence, fourth amino acid sequence and fifth amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ of T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for an edge portion of T08446_PEA_1_P18, comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence encoding for QVSAQLRAGGGGRDAPEAAAQSPCSVPS, corresponding to T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence APQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAP WGPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYP TPSWSLHSEGQTRSYC in T08446_PEA_1_P18.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T11628_PEA_1_P2, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE corresponding to amino acids 1 - 55 of T11628_PEA_1_P2, and a second amino acid sequence being at least 90 % homologous to

MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 1 - 99 of Q8WVH6, which also corresponds to amino acids 56 - 154 of T11628_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T11628_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE of T11628_PEA_1_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T11628_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 56 - 154 of MYG_HUMAN_V1, which also corresponds to amino acids 1 - 99 of T11628_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T11628_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMK ASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQ SKHPGDFGADAQGAMNK corresponding to amino acids 1 - 134 of MYG_HUMAN_V1, which also corresponds to amino acids 1 - 134 of T11628_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence G corresponding to amino acids 135 - 135 of T11628_PEA_1_P7, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T11628_PEA_1_P10, comprising a first amino acid

sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE corresponding to amino acids 1 - 55 of T11628_PEA_1_P10, and a second amino acid sequence being at least 90 % homologous to MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 1 - 99 of Q8WVH6, which also corresponds to amino acids 56 - 154 of T11628_PEA_1_P10, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of T11628_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE of T11628_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P9, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEV corresponding to amino acids 1 - 274 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 274 of R35137_PEA_1_PEA_1_PEA_1_P9, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRAYEAGGGSRAMARPSSPDGPPPPPHLTWPCAGAGSAAAMWRW corresponding to amino acids 275 - 385 of R35137_PEA_1_PEA_1_PEA_1_P9, wherein said

first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRAYEAGGGSRAMARPSSPDGPPPPPHLTWPCAGAGSAAAMWRW in R35137_PEA_1_PEA_1_PEA_1_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVL MEMGPPYAGQQELASFHSTSKGYMGEC corresponding to amino acids 1 - 320 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 320 of R35137_PEA_1_PEA_1_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRTRRVGARGPWPGPPRPMGHPLLRT corresponding to amino acids 321 - 346 of R35137_PEA_1_PEA_1_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRTRRVGARGPWPGPPRPMGHPLLRT in R35137_PEA_1_PEA_1_PEA_1_P8.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQAR corresponding to amino acids 1 - 229 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 229 of R35137_PEA_1_PEA_1_PEA_1_P11, and a second amino acid sequence being at least 90 % homologous to SGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS corresponding to amino acids 455 - 496 of ALAT_HUMAN_V1, which also corresponds to amino acids 230 - 271 of R35137_PEA_1_PEA_1_PEA_1_P11, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of R35137_PEA_1_PEA_1_PEA_1_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise RS, having a structure as follows: a sequence starting from any of amino acid numbers 229-x to 229; and ending at any of amino acid numbers 230+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEV corresponding to amino acids 1 -

274 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 274 of R35137_PEA_1_PEA_1_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRVPRRLCGGGEHGRCSAAADAEADECAAVPAGARTGPAGPGGGQPAR AHRPLLCAVPG corresponding to amino acids 275 - 399 of R35137_PEA_1_PEA_1_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRVPRRLCGGGEHGRCSAAADAEADECAAVPAGARTGPAGPGGGQPAR AHRPLLCAVPG in R35137_PEA_1_PEA_1_PEA_1_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVL MEMGPPYAGQQELASFHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRL CPPVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAELAAKAKLTEQVFNEAPGISCNP VQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGICVVPGSGFGQREGTYH FRMTILPPLEKLRLLLEKLSRFHAKFTLE corresponding to amino acids 1 - 494 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 494 of R35137_PEA_1_PEA_1_PEA_1_P4, and a second amino acid sequence being at least 70%,

optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPGRLWSPLYLLLMPGGVGWGGCWAPASLQVPNKAVWQSDSKKEALAAAWPAPTCL PFLQA corresponding to amino acids 495 - 555 of R35137_PEA_1_PEA_1_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPGRLWSPLYLLLMPGGVGWGGCWAPASLQVPNKAVWQSDSKKEALAAAWPAPTCL PFLQA in R35137_PEA_1_PEA_1_PEA_1_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAGSPCRGLAPGREEQRALHKAGAVGGGVR corresponding to amino acids 1 - 110 of R11723_PEA_1_P6, and a second amino acid sequence being at least 90 % homologous to MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLREGEEDHV RPEVGPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 1 - 112 of Q8IXM0, which also corresponds to amino acids 111 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV

MEQSAGIMYRKSCASSAACLIASAGSPCRGLAPGREEQRALHKAGAVGGGVR of R11723_PEA_1_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 1 - 83 of Q96AC2, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 1 - 83 of Q8N2G4, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL

RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 24 - 106 of BAC85518, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAG corresponding to amino acids 1 - 64 of Q96AC2, which also corresponds to amino acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAG corresponding to amino acids 1 - 64 of Q8N2G4, which also corresponds to amino acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at

least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MWVLG corresponding to amino acids 1 - 5 of R11723_PEA_1_P7, second amino acid sequence being at least 90 % homologous to IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSAG corresponding to amino acids 22 - 80 of BAC85273, which also corresponds to amino acids 6 - 64 of R11723_PEA_1_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLG of R11723_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV

MEQSAG corresponding to amino acids 24 - 87 of BAC85518, which also corresponds to amino acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P13, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q96AC2, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DTKRTNTLLFEMRHFAKQLTT corresponding to amino acids 64 - 84 of R11723_PEA_1_P13, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DTKRTNTLLFEMRHFAKQLTT in R11723_PEA_1_P13.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to

MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q96AC2, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q8N2G4, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MWVLG corresponding to amino acids 1 - 5 of R11723_PEA_1_P10, second amino acid sequence being at least 90 % homologous to IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSA corresponding to amino acids 22 - 79 of BAC85273, which also corresponds to amino acids 6 - 63 of R11723_PEA_1_P10, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLG of R11723_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSA corresponding to amino acids 24 - 86 of BAC85518, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most

preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R16276_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPG corresponding to amino acids 1 - 41 of NOV_HUMAN, which also corresponds to amino acids 1 - 41 of R16276_PEA_1_P7, a bridging amino acid Q corresponding to amino acid 42 of R16276_PEA_1_P7, a second amino acid sequence being at least 90 % homologous to CPATPPTCAPGVRAVLDGCSCCLVCARQRGESCSDLEPCDESSGLYCDRSADPSNQTGI CT corresponding to amino acids 43 - 103 of NOV_HUMAN, which also corresponds to amino acids 43 - 103 of R16276_PEA_1_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GNPAPSAV corresponding to amino acids 104 - 111 of R16276_PEA_1_P7, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R16276_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GNPAPSAV in R16276_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R16276_PEA_1_P7, comprising a first amino acid

sequence being at least 90 % homologous to MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPG corresponding to amino acids 1 - 41 of NOV_HUMAN, which also corresponds to amino acids 1 - 41 of R16276_PEA_1_P7, a bridging amino acid Q corresponding to amino acid 42 of R16276_PEA_1_P7, a second amino acid sequence being at least 90 % homologous to CPATPPTCAPGVRAVLDGCSCCLVCARQRGESCSDLEPCDESSGLYCDRSADPSNQTGICT corresponding to amino acids 43 - 103 of NOV_HUMAN, which also corresponds to amino acids 43 - 103 of R16276_PEA_1_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GNPAPSAV corresponding to amino acids 104 - 111 of R16276_PEA_1_P7, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R16276_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GNPAPSAV in R16276_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVL corresponding to amino acids 1 - 234 of CEA5_HUMAN, which also corresponds to amino acids 1 - 234 of HUMCEA_PEA_1_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence CEYICSSLAQAASPNPQGQRQDFSVPLRFKYTDPQPWTSRLSVTFCPRKTWADQVLTKNRRGGAASVLGGSGSTPYDGRNR corresponding to amino acids 235 - 315 of

HUMCEA_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCEA_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence CEYICSSLAQAASPNPQGQRQDFSVPLRFKYTDPQPWTSRLSVTFCPRKTWADQVLTKN RRGGAASVLGGSGSTPYDGRNR in HUMCEA_PEA_1_P4.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWV NNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDA PTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTC QAHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWV NNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDD PTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQ ANNSASGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVN GQSLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTP IISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFV SNLATGRNNSIVKSITVS corresponding to amino acids 1 - 675 of CEA5_HUMAN, which also corresponds to amino acids 1 - 675 of HUMCEA_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKWLPGASASYSGVESIWFSPKSQEDIFFPSLCSMGTRKSQILS corresponding to amino acids 676 - 719 of HUMCEA_PEA_1_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMCEA_PEA_1_P5, comprising a polypeptide

being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKWLPGASASYSGVESIWFSPKSQEDIFFPSLCSMGTRKSQILS in HUMCEA_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P19, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWV NNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILN corresponding to amino acids 1 - 232 of CEA5_HUMAN, which also corresponds to amino acids 1 - 232 of HUMCEA_PEA_1_P19, and a second amino acid sequence being at least 90 % homologous to VLYGPDTPIISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVALI corresponding to amino acids 589 - 702 of CEA5_HUMAN, which also corresponds to amino acids 233 - 346 of HUMCEA_PEA_1_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HUMCEA_PEA_1_P19, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise NV, having a structure as follows: a sequence starting from any of amino acid numbers 232-x to 232; and ending at any of amino acid numbers 233+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P20, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT

LHVIKSDLVNEEATGQFRVYP corresponding to amino acids 1 - 142 of CEA5_HUMAN, which also corresponds to amino acids 1 - 142 of HUMCEA_PEA_1_P20, and a second amino acid sequence being at least 90 % homologous to ELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLT LFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGANLNLSCHS ASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASG TSPGLSAGATVGIMIGVLVGVALI corresponding to amino acids 499 - 702 of CEA5_HUMAN, which also corresponds to amino acids 143 - 346 of HUMCEA_PEA_1_P20, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HUMCEA_PEA_1_P20, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise PE, having a structure as follows: a sequence starting from any of amino acid numbers 142-x to 142; and ending at any of amino acid numbers 143+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z44808_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ ELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ corresponding to amino acids 1 - 441 of SMO2_HUMAN, which also corresponds to amino acids 1 - 441 of Z44808_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least

85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DAMVVSSRPKATTHRKSRTLSRR corresponding to amino acids 442 - 464 of Z44808_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z44808_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DAMVVSSRPKATTHRKSRTLSRR in Z44808_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z44808_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ ELMGCLGVAKEDGKADTKKRH corresponding to amino acids 1 - 428 of SMO2_HUMAN, which also corresponds to amino acids 1 - 428 of Z44808_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RSKRNL corresponding to amino acids 429 - 434 of Z44808_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z44808_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RSKRNL in Z44808_PEA_1_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z44808_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ ELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ corresponding to amino acids 1 - 441 of SMO2_HUMAN, which also corresponds to amino acids 1 - 441 of Z44808_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LLWLRGKVSFYCF corresponding to amino acids 442 - 454 of Z44808_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z44808_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LLWLRGKVSFYCF in Z44808_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z44808_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKT corresponding to amino acids 1 - 170 of SMO2_HUMAN, which also corresponds to amino acids 1 - 170 of Z44808_PEA_1_P11, and a second amino acid sequence being at least 90 % homologous to

DIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGL
YKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQ
GCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLD
KNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQELMGCLGVAKE
DGKADTKKRHTPRGHAESTSNRQPRKQG corresponding to amino acids 188 - 446 of SMO2_HUMAN, which also corresponds to amino acids 171 - 429 of Z44808_PEA_1_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of Z44808_PEA_1_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TD, having a structure as follows: a sequence starting from any of amino acid numbers 170-x to -170; and ending at any of amino acid numbers 171+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H61775_P16, comprising a first amino acid sequence being at least 90 % homologous to
MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL
RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 11 - 93 of Q9P2J2, which also corresponds to amino acids 1 - 83 of H61775_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence
DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW
RSSCSVTLQV corresponding to amino acids 84 - 152 of H61775_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of H61775_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV in H61775_P16.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H61775_P16, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 1 - 83 of AAQ88495, which also corresponds to amino acids 1 - 83 of H61775_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV corresponding to amino acids 84 - 152 of H61775_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of H61775_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV in H61775_P16.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H61775_P17, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 11 - 93 of Q9P2J2, which also corresponds to amino acids 1 - 83 of H61775_P17.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H61775_P17, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 1 - 83 of AAQ88495, which also corresponds to amino acids 1 - 83 of H61775_P17.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M85491_PEA_1_P13, comprising a first amino acid sequence being at least 90 % homologous to MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIR TYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYY EADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGGRVMKINTEVRSFGPVSRSGF YLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAEEVD VPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPIN SRTTSEGATNCVCRNGYYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDSG GREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQLGLTEPRIYISDLLAHTQYTFEIQAV NGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVDSITLSWSQPDQPNGVILDYEL QYYEK corresponding to amino acids 1 - 476 of EPB2_HUMAN, which also corresponds to amino acids 1 - 476 of M85491_PEA_1_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPIGWVLSPSPTSLRAPLPG corresponding to amino acids 477 - 496 of M85491_PEA_1_P13, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M85491_PEA_1_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPIGWVLSPSPTSLRAPLPG in M85491_PEA_1_P13.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for M85491_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIR TYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYY EADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGGRVMKINTEVRSFGPVSRSGF YLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAEEVD VPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCR corresponding to amino acids 1 -

270 of EPB2_HUMAN, which also corresponds to amino acids 1 - 270 of M85491_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ERQDLTMLSRLVLNSWPQMILPPQPPKVLEL corresponding to amino acids 271 - 301 of M85491_PEA_1_P14, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of M85491_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ERQDLTMLSRLVLNSWPQMILPPQPPKVLEL in M85491_PEA_1_P14.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P6, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKG corresponding to amino acids 1 - 276 of VTNC_HUMAN, which also corresponds to amino acids 1 - 276 of T39971_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TQGVVGD corresponding to amino acids 277 - 283 of T39971_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of T39971_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TQGVVGD in T39971_P6.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P9, comprising a first amino acid sequence being at least 90 % homologous to

MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC
KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV
LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR
GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV
LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE
CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRT corresponding to amino acids 1 - 325 of VTNC_HUMAN, which also corresponds to amino acids 1 - 325 of T39971_P9, and a second amino acid sequence being at least 90 % homologous to

SGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRATWLSLFSSEESNLGA
NNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLRTRRVDTVDPPYPRSIAQYWLGC
PAPGHL corresponding to amino acids 357 - 478 of VTNC_HUMAN, which also corresponds to amino acids 326 - 447 of T39971_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of T39971_P9, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TS, having a structure as follows: a sequence starting from any of amino acid numbers 325-x to 325; and ending at any of amino acid numbers 326 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P11, comprising a first amino acid sequence being at least 90 % homologous to

MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC
KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV
LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR
GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV

LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRTS corresponding to amino acids 1 - 326 of VTNC_HUMAN, which also corresponds to amino acids 1 - 326 of T39971_P11, and a second amino acid sequence being at least 90 % homologous to DKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL corresponding to amino acids 442 - 478 of VTNC_HUMAN, which also corresponds to amino acids 327 - 363 of T39971_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of T39971_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise SD, having a structure as follows: a sequence starting from any of amino acid numbers 326-x to 326; and ending at any of amino acid numbers 327 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P11, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRTS corresponding to amino acids 1 - 326 of Q9BSH7, which also corresponds to amino acids 1 - 326 of T39971_P11, and a second amino acid sequence being at least 90 % homologous to DKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL corresponding to amino acids 442 - 478 of Q9BSH7, which also corresponds to amino acids 327 - 363 of T39971_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of T39971_P11, comprising a

polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise SD, having a structure as follows: a sequence starting from any of amino acid numbers 326-x to 326; and ending at any of amino acid numbers 327 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P12, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFK corresponding to amino acids 1 - 223 of VTNC_HUMAN, which also corresponds to amino acids 1 - 223 of T39971_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPGAVGQGRKHLGRV corresponding to amino acids 224 - 238 of T39971_P12, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of T39971_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPGAVGQGRKHLGRV in T39971_P12.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for T39971_P12, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFK corresponding to

amino acids 1 - 223 of Q9BSH7, which also corresponds to amino acids 1 - 223 of T39971_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPGAVGQGRKHLGRV corresponding to amino acids 224 - 238 of T39971_P12, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of T39971_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPGAVGQGRKHLGRV in T39971_P12.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to

MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW
QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR
NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ
FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD
SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP
GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFL
VERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGK
LRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQ
GTPKYKPRFVHTRQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQ
ASSGGNRGRMLADSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYI
DKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKE
AAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWN

corresponding to amino acids 1 - 761 of SUL1_HUMAN, which also corresponds to amino acids 1 - 761 of Z21368_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

PHKYSAHGRTRHFESATRTTNGAQKLSRI corresponding to amino acids 762 - 790 of Z21368_PEA_1_P2, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z21368_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PHKYSAHGRTRHFESATRTTNGAQKLSRI in Z21368_PEA_1_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVEL corresponding to amino acids 1 - 57 of Q7Z2W2, which also corresponds to amino acids 1 - 57 of Z21368_PEA_1_P5, second bridging amino acid sequence comprising A, and a third amino acid sequence being at least 90 % homologous to FFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITN ESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNM DKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYT ADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDT PPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHL PKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLY ARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFE GEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPT TVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKR RKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKER KEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNE THNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCN PRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 139 - 871 of Q7Z2W2, which also corresponds to amino acids 59 - 791 of Z21368_PEA_1_P5, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for an edge portion of Z21368_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise LAF having a structure as follows (numbering according to Z21368_PEA_1_P5): a sequence starting from any of amino acid numbers 57-x to 57; and ending at any of amino acid numbers 59 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELAFF GKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNES INYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNMDK HWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYTAD HGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPP DVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLP KYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYA RGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFEGE IYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTV RVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRK PEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKE KRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETH NFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLME corresponding to amino acids 1 - 751 of Z21368_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to LRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 1 - 40 of AAH12997, which also corresponds to amino acids 752 - 791 of Z21368_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of Z21368_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELAFF GKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNES INYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNMDK HWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYTAD HGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPP DVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLP KYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYA RGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFEGE IYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTV RVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRK PEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKE KRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETH NFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLME of Z21368_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVEL corresponding to amino acids 1 - 57 of SUL1_HUMAN, which also corresponds to amino acids 1 - 57 of Z21368_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to AFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLIT NESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPN MDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYII YTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGL DTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSN HLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRN

LYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVE
FEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGP
PTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHL
KRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRK
KERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRT
VNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYK
QCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 138 - 871 of SUL1_HUMAN, which also corresponds to amino acids 58 - 791 of Z21368_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of Z21368_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise LA, having a structure as follows: a sequence starting from any of amino acid numbers 57-x to 57; and ending at any of amino acid numbers 58 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P15, comprising a first amino acid sequence being at least 90 % homologous to
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW
QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR
NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ
FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD
SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP
GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFL
VERG corresponding to amino acids 1 - 416 of SUL1_HUMAN, which also corresponds to amino acids 1 - 416 of Z21368_PEA_1_P15.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P16, comprising a first amino acid

sequence being at least 90 % homologous to
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW
QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR
NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ
FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD
SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP
GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNR corresponding to amino acids 1 - 397 of SUL1_HUMAN, which also corresponds to amino acids 1 - 397 of Z21368_PEA_1_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence CVIVPPLSQPQIH corresponding to amino acids 398 - 410 of Z21368_PEA_1_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z21368_PEA_1_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence CVIVPPLSQPQIH in Z21368_PEA_1_P16.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P22, comprising a first amino acid sequence being at least 90 % homologous to
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW
QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR
NGIKEKHGFDYAK corresponding to amino acids 1 - 188 of SUL1_HUMAN, which also corresponds to amino acids 1 - 188 of Z21368_PEA_1_P22, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ARYDGDQPRCAPRPRGLSPTVF corresponding to amino acids 189 - 210 of

Z21368_PEA_1_P22, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z21368_PEA_1_P22, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ARYDGDQPRCAPRPRGLSPTVF in Z21368_PEA_1_P22.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P23, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRT corresponding to amino acids 1 - 137 of Q7Z2W2, which also corresponds to amino acids 1 - 137 of Z21368_PEA_1_P23, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GLLHRLNH corresponding to amino acids 138 - 145 of Z21368_PEA_1_P23, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z21368_PEA_1_P23, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GLLHRLNH in Z21368_PEA_1_P23.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z21368_PEA_1_P23, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRT corresponding to amino acids 1 - 137 of SUL1_HUMAN, which also corresponds to amino acids 1 - 137 of Z21368_PEA_1_P23, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more

preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GLLHRLNH corresponding to amino acids 138 - 145 of Z21368_PEA_1_P23, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z21368_PEA_1_P23, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GLLHRLNH in Z21368_PEA_1_P23.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMGRP5E_P4, comprising a first amino acid sequence being at least 90 % homologous to MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTG ESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSED SSNFKDVGSKGK corresponding to amino acids 1 - 127 of GRP_HUMAN, which also corresponds to amino acids 1 - 127 of HUMGRP5E_P4, and a second amino acid sequence being at least 90 % homologous to GSQREGRNPQLNQQ corresponding to amino acids 135 - 148 of GRP_HUMAN, which also corresponds to amino acids 128 - 141 of HUMGRP5E_P4, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of HUMGRP5E_P4, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KG, having a structure as follows: a sequence starting from any of amino acid numbers 127-x to 127; and ending at any of amino acid numbers 128 + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMGRP5E_P5, comprising a first amino acid sequence being at least 90 % homologous to MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTG ESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSED

SSNFKDVGSKGK corresponding to amino acids 1 - 127 of GRP_HUMAN, which also corresponds to amino acids 1 - 127 of HUMGRP5E_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DSLLQVLNVKEGTPS corresponding to amino acids 128 - 142 of HUMGRP5E_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of HUMGRP5E_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DSLLQVLNVKEGTPS in HUMGRP5E_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for D56406_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAHVPSWKMT LLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEAMLTIYQLHKICHSRAF QHWE corresponding to amino acids 1 - 120 of NEUT_HUMAN, which also corresponds to amino acids 1 - 120 of D56406_PEA_1_P2, second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ARWLTPVIPALWEAETGGSRGQEMETIPANT corresponding to amino acids 121 - 151 of D56406_PEA_1_P2, and a third amino acid sequence being at least 90 % homologous to LIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 121 - 170 of NEUT_HUMAN, which also corresponds to amino acids 152 - 201 of D56406_PEA_1_P2, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for an edge portion of D56406_PEA_1_P2, comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the

sequence encoding for ARWLTPVIPALWEAETGGSRGQEMETIPANT, corresponding to D56406_PEA_1_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for D56406_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLC corresponding to amino acids 1 - 23 of NEUT_HUMAN, which also corresponds to amino acids 1 - 23 of D56406_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to SEEEMKALEADFLTNMHTSKISKAHVPSWKMTLLNVCSLVNNLNSPAEETGEVHEEEL VARRKLPTALDGFSLEAMLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKR KIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 26 - 170 of NEUT_HUMAN, which also corresponds to amino acids 24 - 168 of D56406_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of D56406_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise CS, having a structure as follows: a sequence starting from any of amino acid numbers 23-x to 24; and ending at any of amino acid numbers + ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for D56406_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLCSDS EEEMKALEADFLTNMHTSK corresponding to amino acids 1 - 45 of NEUT_HUMAN, which also corresponds to amino acids 1 - 45 of D56406_PEA_1_P6, and a second amino acid sequence being at least 90 % homologous to LIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 121 - 170 of NEUT_HUMAN, which also corresponds to amino acids 46 - 95 of D56406_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for an edge portion of D56406_PEA_1_P6, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KL, having a structure as follows: a sequence starting from any of amino acid numbers 45-x to 46; and ending at any of amino acid numbers 46+ ((n-2) - x), in which x varies from 0 to n-2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for F05068_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MKLVSVALMYLGSLAFLGADTARLDVASEFRKK corresponding to amino acids 1 - 33 of ADML_HUMAN, which also corresponds to amino acids 1 - 33 of F05068_PEA_1_P7.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for F05068_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSSSYPTGLA DVKAGPAQTLIRPQDMKGASRSPED corresponding to amino acids 1 - 82 of ADML_HUMAN, which also corresponds to amino acids 1 - 82 of F05068_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence R corresponding to amino acids 83 - 83 of F05068_PEA_1_P8, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H14624_P15, comprising a first amino acid sequence being at least 90 % homologous to MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLCHGIEYQNMR LPNLLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFAPVCLDDLDETIQPCHSLC VQVKDRCAPVMSAFGFPWPDMLECDRFPQDNDLCIPLASSDHLLPATEE corresponding to amino acids 1 - 167 of Q9HAP5, which also corresponds to amino acids 1 - 167 of H14624_P15, and a second amino acid sequence being at least 70%, optionally at least 80%,

preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKPSLLLPHSLLG corresponding to amino acids 168 - 180 of H14624_P15, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of H14624_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKPSLLLPHSLLG in H14624_P15.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H38804_PEA_1_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ corresponding to amino acids 1 - 57 of H38804_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGA VLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIRCVEYCPEVNVMVTG SWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQ QRRESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENN IEQIYPVNAISFHNIHNTFATGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTL AIASSYMYEMDDTEHPEDGIFIRQVTDAETKPK corresponding to amino acids 1 - 324 of BUB3_HUMAN, which also corresponds to amino acids 58 - 381 of H38804_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of H38804_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ of H38804_PEA_1_P5.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for H38804_PEA_1_P17, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ corresponding to amino acids 1 - 57 of H38804_PEA_1_P17, and a second amino acid sequence being at least 90 % homologous to MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGA VLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIRCVEYCPEVNVMVTG SWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQ QRRESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENN IEQIYPVNAISFHNIHNTFATGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTL AIASSYMYEMDDTEHPEDGIFIRQVTDAETKPKSPCT corresponding to amino acids 1 - 328 of BUB3_HUMAN, which also corresponds to amino acids 58 - 385 of H38804_PEA_1_P17, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of H38804_PEA_1_P17, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ of H38804_PEA_1_P17.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HSENA78_P2, comprising a first amino acid sequence being at least 90 % homologous to MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCVCLQTTQGVHP KMISNLQVFAIGPQCSKVEVV corresponding to amino acids 1 - 81 of SZ05_HUMAN, which also corresponds to amino acids 1 - 81 of HSENA78_P2.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at

least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA WESGMKRHRAACASASINV corresponding to amino acids 151 - 461 of DCOR_HUMAN, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA WESGMKRHRAACASASINV corresponding to amino acids 40 - 350 of AAA59968, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA WESGMKRHRAACASASINV corresponding to amino acids 86 - 396 of AAH14562, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R00299_P3, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV corresponding to amino acids 1 - 44 of R00299_P3, second amino acid sequence being at least 90 % homologous to

SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNRNLRKGPSGLA DEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRITLEEYRNV corresponding to amino acids 74 - 191 of Q9NWT9, which also corresponds to amino acids 45 - 162 of R00299_P3, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VEELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIE TKMHVRFLNMETMALCH corresponding to amino acids 163 - 238 of R00299_P3, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV of R00299_P3.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VEELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIE TKMHVRFLNMETMALCH in R00299_P3.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for R00299_P3, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV corresponding to amino acids 1 - 44 of R00299_P3, and a second amino acid sequence being at least 90 % homologous to SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNRNLRKGPSGLA DEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRITLEEYRNVVE ELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIETK MHVRFLNMETMALCH corresponding to amino acids 21 - 214 of TESC_HUMAN, which

also corresponds to amino acids 45 - 238 of R00299_P3, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a head of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV of R00299_P3.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for W60282_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTA AHCLKP corresponding to amino acids 1 - 66 of Q8IXD7, which also corresponds to amino acids 1 - 66 of W60282_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TPASHLAMRQHHHH corresponding to amino acids 67 - 80 of W60282_PEA_1_P14, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of W60282_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TPASHLAMRQHHHH in W60282_PEA_1_P14.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 1 - 95 of SZ14_HUMAN, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to

amino acids 96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 13 - 107 of Q9NS21, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to amino acids 96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 13 - 107 of AAQ89265, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide

having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to amino acids 96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

According to preferred embodiments of the present invention, there is provided an isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

According to preferred embodiments of the present invention, there is provided an antibody capable of specifically binding to an epitope of an amino acid sequences.

Optionally the amino acid sequence corresponds to a bridge, edge portion, tail, head or insertion.

Optionally the antibody is capable of differentiating between a splice variant having said epitope and a corresponding known protein.

According to preferred embodiments of the present invention, there is provided a kit for detecting lung cancer, comprising a kit detecting overexpression of a splice variant according to any of the above claims.

Optionally the kit comprises a NAT-based technology.

Optionally the kit further comprises at least one primer pair capable of selectively hybridizing to a nucleic acid sequence according to any of the above claims.

Optionally the kit further comprises at least one oligonucleotide capable of selectively hybridizing to a nucleic acid sequence according to any of the above claims.

Optionally the kit comprises an antibody according to any of the above claims.

Optionally the kit further comprises at least one reagent for performing an ELISA or a Western blot.

According to preferred embodiments of the present invention, there is provided a method for detecting lung cancer, comprising detecting overexpression of a splice variant according to any of the above claims.

Optionally the detecting overexpression is performed with a NAT-based technology.

Optionally detecting overexpression is performed with an immunoassay.

Optionally the immunoassay comprises an antibody according to any of the above claims.

According to preferred embodiments of the present invention, there is provided a biomarker capable of detecting lung cancer, comprising any of the above nucleic acid sequences or a fragment thereof, or any of the above amino acid sequences or a fragment thereof.

According to preferred embodiments of the present invention, there is provided a method for screening for lung cancer, comprising detecting lung cancer cells with a biomarker or an antibody or a method or assay according to any of the above claims.

According to preferred embodiments of the present invention, there is provided a method for diagnosing lung cancer, comprising detecting lung cancer cells with a biomarker or an antibody or a method or assay according to any of the above claims.

According to preferred embodiments of the present invention, there is provided a method for monitoring disease progression and/or treatment efficacy and/or relapse of lung cancer, comprising detecting lung cancer cells with a biomarker or an antibody or a method or assay according to any of the above claims.

According to preferred embodiments of the present invention, there is provided a method of selecting a therapy for lung cancer, comprising detecting lung cancer cells with a biomarker or an antibody or a method or assay according to any of the above claims and selecting a therapy according to said detection.

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). All of these are hereby incorporated by reference as if fully set forth herein. As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is schematic summary of cancer biomarkers selection engine and the wet validation stages.

Figure 2. Schematic illustration, depicting grouping of transcripts of a given contig based on presence or absence of unique sequence regions.

Figure 3 is schematic summary of quantitative real-time PCR analysis.

Figure 4 is schematic presentation of the oligonucleotide based microarray fabrication.

Figure 5 is schematic summary of the oligonucleotide based microarray experimental flow.

Figure 6 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster H61775, demonstrating overexpression in brain malignant tumors and a mixture of malignant tumors from different tissues.

Figure 7 is a histogram showing expression of transcripts of variants of the immunoglobulin superfamily, member 9,H61775 transcripts, which are detectable by amplicon as depicted in sequence name H61775seg8, in normal and cancerous lung tissues.

Figure 8 is a histogram showing expression of immunoglobulin superfamily, member 9, H61775 transcripts, which are detectable by amplicon as depicted in sequence name H61775seg8, in different normal tissues.

Figure 9 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster M85491, demonstrating overexpression in epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Figure 10 is a histogram showing over expression of the above-indicated Ephrin type-B receptor 2 precursor M85491 transcripts, which are detectable by amplicon as depicted in sequence name M85491seg24, in cancerous lung samples relative to the normal samples.

Figure 11 is a histogram showing the expression of Ephrin type-B receptor 2 precursor (Tyrosine-protein kinase receptor EPH-3) M85491 transcripts which are detectable by amplicon as depicted in sequence name M85491seg24 in different normal tissues.

Figure 12 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster T39971, demonstrating overexpression in liver cancer, lung malignant tumors and pancreas carcinoma.

Figure 13 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster Z21368, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Figure 14 is a histogram showing over expression of the Extracellular sulfatase Sulf-1 Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368junc17-21, in cancerous lung samples relative to the normal samples.

Figure 15 is a histogram showing the expression of Extracellular sulfatase Sulf-1 Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368 junc17-21, in different normal tissues.

. Figure 16 is a histogram showing over expression of the SUL1_HUMAN - Extracellular sulfatase Sulf-1, Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368seg39, in cancerous lung samples relative to the normal samples.

Figure 17 is a histogram showing expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1, Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368seg39, in different normal tissues.

Figure18 is a histogram showing the expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor (Secreted modular calcium-binding protein 2) (SMOC-2) (Smooth muscle-associated protein 2) Z44808 transcripts which are detectable by amplicon as depicted in sequence name Z44808 junc8-11 in different normal tissues.

Figure 19 is a histogram showing over expression of the gastrin-releasing peptide (HUMGRP5E) transcripts, which are detectable by amplicon as depicted in sequence name HUMGRP5Ejunc3-7, in several cancerous lung samples relative to the normal samples.

Figure 20 is a histogram showing the expression of gastrin-releasing peptide (HUMGRP5E) transcripts, which are detectable by amplicon as depicted in sequence name HUMGRP5Ejunc3-7, in different normal tissues.

Figure 21 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster F05068, demonstrating overexpression in uterine malignancies.

Figure 22 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster H14624, demonstrating overexpression in colorectal cancer, epithelial malignant tumors, a mixture of malignant tumors from different tissues, lung malignant tumors and pancreas carcinoma.

Figure 23 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster H38804, demonstrating overexpression in transitional cell carcinoma, brain malignant tumors, a mixture of malignant tumors from different tissues and gastric carcinoma.

Figure 24 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HSENA78, demonstrating overexpression in epithelial malignant tumors and lung malignant tumors.

Figure 25 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMODCA, demonstrating overexpression in : brain malignant tumors, colorectal cancer, epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Figure 26 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster R00299, demonstrating overexpression in lung malignant tumors.

Figure 27 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster Z41644, demonstrating overexpression in lung malignant tumors, breast malignant tumors and pancreas carcinoma.

Figure 28 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster Z44808, demonstrating overexpression in colorectal cancer, lung cancer and pancreas carcinoma.

Figure 29 is a histogram showing over expression of the SMO2_HUMAN SPARC related modular calcium-binding protein 2 Z44808 transcripts, which are detectable by amplicon as depicted in sequence name Z44808junc8-11, in cancerous lung samples relative to the normal samples.

Figure 30 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster AA161187, demonstrating overexpression in brain malignant tumors, epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Figure 31 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster AA161187, demonstrating overexpression in brain malignant tumors and a mixture of malignant tumors from different tissues.

Figure 32 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMCA1XIA, demonstrating overexpression in bone malignant tumors, epithelial malignant tumors, a mixture of malignant tumors from different tissues and lung malignant tumors.

Figure 33 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMCEA, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Figure 34 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster R35137, demonstrating overexpression in hepatocellular carcinoma.

Figure 35 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster Z25299, demonstrating overexpression in brain malignant tumors, a mixture of malignant tumors from different tissues and ovarian carcinoma.

Figure 36 is a histogram showing down regulation of the Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts, which are detectable by amplicon as depicted in sequence name Z25299 junc13-14-21, in cancerous lung samples relative to the normal samples.

Figure 37 is a histogram showing down regulation of the Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts, which are detectable by amplicon as depicted in sequence name Z25299 seg20, in cancerous lung samples relative to the normal samples.

Figure 38 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HSSTROL3, demonstrating overexpression in transitional cell carcinoma, epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Figure 39 is a histogram showing over expression of the Stromelysin-3 HSSTROL3 transcripts, which are detectable by amplicon as depicted in sequence name HSSTROL3 seg24, in cancerous lung samples relative to the normal samples.

Figure 40 is a histogram showing the expression of Stromelysin-3 HSSTROL3 transcripts, which are detectable by amplicon as depicted in sequence name HSSTROL3 seg24, in different normal tissues.

Figure 41 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMTREFAC, demonstrating overexpression in a mixture of malignant tumors from different tissues, breast malignant tumors, pancreas carcinoma and prostate cancer.

Figure 42 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HSS100PCB, demonstrating overexpression in a mixture of malignant tumors from different tissues.

Figure 43 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HSU33147, demonstrating overexpression in a mixture of malignant tumors from different tissues.

Figure 44 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster R20779, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues and lung malignant tumors.

Figure 45 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster R38144, demonstrating overexpression in epithelial malignant tumors, lung malignant tumors, skin malignancies and gastric carcinoma.

Figure 46 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMOSTRO, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues, lung malignant tumors, breast malignant tumors, ovarian carcinoma and skin malignancies.

Figure 47 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster HUMOSTRO, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues and kidney malignant tumors.

Figure 48 is a histogram showing over expression of the R11723 transcripts, which are detectable by amplicon as depicted in sequence name R11723 seg13, in cancerous lung samples relative to the normal samples.

Figure 49 is a histogram showing the expression of R11723 transcripts which are detectable by amplicon as depicted in sequence name R11723seg13 in different normal tissues.

Figure 50 is a histogram showing over expression of the R11723 transcripts, which are detectable by amplicon as depicted in sequence name R11723 junc11-18 in cancerous lung samples relative to the normal samples.

Figure 51 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster R16276, demonstrating overexpression in: lung malignant tumors.

Figures 52-53 are histograms, showing differential expression of the 6 sequences H61775seg8, HUMGRP5E junc3-7, M85491Seg24, Z21368 junc17-21, HSSTROL3seg24 and Z25299seg20 in cancerous lung samples relative to the normal samples.

Figure 54a is a histogram showing the relative expression of trophinin associated protein (tastin)) [T86235] variants (e.g., variant no. 23-26, 31, 32) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1480.

Figure 54b is a histogram showing the relative expression of trophinin associated protein (tastin)) [T86235] variants (e.g., variant no. 8-10, 22, 23, 26,27, 29-31, 33) in normal and tumor derived lung samples as determined micro-array analysis using oligos detailed in SEQ ID NO: 1512-1514.

Figure 55 is a histogram showing the relative expression of Homeo box C10 (HOXC10) [N31842] variants (e.g., variant no. 3) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1517.

Figures 56a-b are histograms showing on two different scales the relative expression of Nucleolar protein 4 (NOL4) [T06014] variants (e.g., variant no. 3, 11 and 12) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1529. Figure 56a shows the results on scale:0-1200. Figure 56b shows the results on scale:0-24.

Figures 57a-b is a histogram showing on two different scales the relative expression of Nucleolar protein 4 (NOL4) [T06014] variants (e.g., variant no. 3, 11 and 12) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1532. Figure 57a shows the results on scale:0-2000. Figure 57b shows the results on scale:0-42.

Figure 58 is a histogram showing the relative expression of AA281370 variants (e.g., variant no. 0, 1, 4 and 5) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1558.

Figure 59 is a histogram showing the relative expression of Sulfatase 1 (SULF1)-[Z21368] variants (e.g., variant no. 13 and 14) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1574.

Figure 60 is a histogram showing the relative expression of SRY (sex determining region Y)-box 2 (SOX2))-[HUMHMGBOX] variants (e.g., variant no. 0) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1594.

Figure 61 is a histogram showing the relative expression of Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome) (PKP1) -[HSB6PR] variants (e.g., variant no. 0, 5 and 6) in normal and tumor derived lung samples as determined by real time PCR using primers for SEQ ID NO: 1600.

Figure 62 is a histogram showing the relative expression of transcripts detectable by SEQ ID NOs: 1480, 1517, 1529, 1532, 1558, 1574, 1594, 1600, 1616, 1619, 1622, 1625 in normal and tumor derived lung samples as determined by real time PCR.

Figure 63 is an amino acid sequence alignment, using NCBI BLAST default parameters, demonstrating similarity between the AA281370 lung cancer biomarker if the present invention to WD40 domains of various proteins involved in MAPK signal trunsduction pathway. Figure 63a: amino acids at positions 40-790 of AA281370 polypeptide SEQ ID NO: 99 has 75% homology to mouse Mapkbp1 protein (gi|47124622). Figure 63b: amino acids at positions 40-886 of the AA281370 polypeptide SEQ ID NO: 99 has 70% homology to rat JNK-binding protein JNKBP1 (gi|34856717).

Figure 64 is a histogram showing over expression of the Homo sapiens protease, serine, 21 (testisin) (PRSS21) AA161187 transcripts, which are detectable by amplicon as depicted in sequence name AA161187 seg25, in cancerous lung samples relative to the normal samples.

Figure 65 is a histogram showing over expression of the protein tyrosine phosphatase, receptor type, S (PTPRS) M62069 transcripts, which are detectable by amplicon as depicted in sequence name M62069 seg19, in cancerous lung samples relative to the normal samples.

Figure 66 is a histogram showing over expression of the protein tyrosine phosphatase, receptor type, S (PTPRS) M62069 transcripts, which are detectable by amplicon as depicted in sequence name M62069 seg29, in cancerous lung samples relative to the normal samples.

Figure 67 is a histogram showing over expression of the above-indicated Homo sapiens collagen, type XI, alpha 1 (COL11A1) transcripts which are detectable by amplicon as depicted in sequence name HUMCA1X1A seg55 in cancerous lung samples relative to the normal samples.

Figure 68 is a histogram showing down regulation of the Homo sapiens secretory leukocyte protease inhibitor (antileukoproteinase) (SLPI) Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299 seg23 in cancerous lung samples relative to the normal samples.

Figure 69 is a histogram showing the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg20 in different normal tissues.

Figure 70 is a histogram showing the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg23 in different normal tissues.

Figure 71 is a histogram showing over expression of the Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 seg20-2 in cancerous lung samples relative to the normal samples.

Figure 72 is a histogram showing over expression of the Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 junc21-27 in cancerous lung samples relative to the normal samples.

Figure 73 is a histogram showing the expression of R11723 transcripts, which were detected by amplicon as depicted in the sequence name R11723 junc11-18 in different normal tissues.

Figure 74 is a histogram showing over expression of the Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 junc24-27F1R3 in cancerous lung samples relative to the normal samples.

Figure 75 is a histogram showing the expression of the Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as

depicted in sequence name H53626 seg25 in cancerous lung samples relative to the normal samples.

Figure 76 is a histogram showing Cancer and cell-line vs. normal tissue expression for Cluster H53626, demonstrating overexpression in epithelial malignant tumors, a mixture of malignant tumors from different tissues and myosarcoma.

Figure 77 is a histogram showing the expression of *of* Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 seg25 in different normal tissues.

Figure 78 is a histogram showing the expression of *of* Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 *junc24-27F1R3* in different normal tissues.

Figure 79 shows PSEC R11723_PEA_1_T5 PCR product; Lane 1: PCR product; and Lane 2: Low DNA Mass Ladder MW marker (Invitrogen Cat# 10068-013).

Figure 80: PSEC R11723_PEA_1 T5 PCR product sequence; In Red- PSEC Forward primer; In Blue- PSEC Reverse complementary sequence; and Highlighted sequence- PSEC variant R11723_PEA_1 T5 ORF.

Figure 81- PRSEC PCR product digested with NheI and HindIII; Lane 1- PRSET PCR product; Lane 2- Fermentas GeneRuler 1Kb DNA Ladder #SM0313.

Figure 82 shows a plasmid map of His PSEC T5 pRSETA.

Figure 83: Protein sequence of PSEC variant R11723_PEA_1 T5; In red- 6His tag; In blue- PSEC.

Figure 84 shows the DNA sequence of HisPSEC T5 pRSETA; bold- HisPSEC T5 open reading frame; Italic- flanking DNA sequence which was verified by sequence analysis.

Figure 85 shows Western blot analysis of recombinant HisPSEC variant R11723_PEA_1 T5; lane 1: molecular weight marker (ProSieve color, Cambrex, Cat #50550); lane 2: HisPSEC T5 pRSETA T0; lane 3: His HisPSEC T5 pRSETA T3; lane 4 :His HisPSEC T5 pRSETA To.n; lane 5: pRSET empty vector T0 (negative control); lane 6: pRSET empty vector T3 (negative control); lane 7: pRSET empty vector To.n (negative control); and lane 8: His positive control protein (HisTroponinT7 pRSETA T3).

DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention is of novel markers for lung cancer that are both sensitive and accurate. Furthermore, at least certain of these markers are able to distinguish between various types of lung cancer, such as small cell carcinoma; large cell carcinoma; squamous cell carcinoma; and adenocarcinoma, alone or in combination. These markers are differentially expressed, and preferably overexpressed, in lung cancer specifically, as opposed to normal lung tissue. The measurement of these markers, alone or in combination, in patient samples provides information that the diagnostician can correlate with a probable diagnosis of lung cancer. The markers of the present invention, alone or in combination, show a high degree of differential detection between lung cancer and non-cancerous states. The markers of the present invention, alone or in combination, can be used for prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. For example, optionally and preferably, these markers may be used for staging lung cancer and/or monitoring the progression of the disease. Furthermore, the markers of the present invention, alone or in combination, can be used for detection of the source of metastasis found in anatomical places other than lung. Also, one or more of the markers may optionally be used in combination with one or more other lung cancer markers (other than those described herein). According to an optional embodiment of the present invention, such a combination may be used to differentiate between various types of lung cancer, such as small cell carcinoma; large cell carcinoma; squamous cell carcinoma; and adenocarcinoma. Furthermore, the markers of the present invention, alone or in combination, can be used for detection of other types of tumors by elimination (for example, for such detection of carcinoid tumors, which are 5% of lung cancers).

The markers of the present invention, alone or in combination, can be used for prognosis, prediction, screening, early diagnosis, staging, therapy selection and treatment monitoring of lung cancer. For example, optionally and preferably, these markers may be used for staging lung cancer and/or monitoring the progression of the disease. Furthermore, the markers of the present invention, alone or in combination, can be used for detection of the source of metastasis found in anatomical places other then lung. Also, one or more of the markers may optionally be used in combination with one or more other lung cancer markers (other than those described herein).

Biomolecular sequences (amino acid and/or nucleic acid sequences) uncovered using the methodology of the present invention and described herein can be efficiently utilized as tissue or pathological markers and/or as drugs or drug targets for treating or preventing a disease.

These markers are specifically released to the bloodstream under conditions of lung cancer, and/or are otherwise expressed at a much higher level and/or specifically expressed in lung cancer tissue or cells. The measurement of these markers, alone or in combination, in patient samples provides information that the diagnostician can correlate with a probable diagnosis of lung cancer.

The present invention therefore also relates to diagnostic assays for lung cancer and/or an indicative condition, and methods of use of such markers for detection of lung cancer and/or an indicative condition, optionally and preferably in a sample taken from a subject (patient), which is more preferably some type of blood sample.

In another embodiment, the present invention relates to bridges, tails, heads and/or insertions, and/or analogs, homologs and derivatives of such peptides. Such bridges, tails, heads and/or insertions are described in greater detail below with regard to the Examples.

As used herein a "tail" refers to a peptide sequence at the end of an amino acid sequence that is unique to a splice variant according to the present invention. Therefore, a splice variant having such a tail may optionally be considered as a chimera, in that at least a first portion of the splice variant is typically highly homologous (often 100% identical) to a portion of the corresponding known protein, while at least a second portion of the variant comprises the tail.

As used herein a "head" refers to a peptide sequence at the beginning of an amino acid sequence that is unique to a splice variant according to the present invention. Therefore, a splice variant having such a head may optionally be considered as a chimera, in that at least a first portion of the splice variant comprises the head, while at least a second portion is typically highly homologous (often 100% identical) to a portion of the corresponding known protein.

As used herein "an edge portion" refers to a connection between two portions of a splice variant according to the present invention that were not joined in the wild type or known protein. An edge may optionally arise due to a join between the above "known protein" portion of a variant and the tail, for example, and/or may occur if an internal portion of the wild type sequence is no longer present, such that two portions of the sequence are now contiguous in the splice variant that were not contiguous in the known protein. A "bridge" may optionally be an

edge portion as described above, but may also include a join between a head and a "known protein" portion of a variant, or a join between a tail and a "known protein" portion of a variant, or a join between an insertion and a "known protein" portion of a variant.

Optionally and preferably, a bridge between a tail or a head or a unique insertion, and a "known protein" portion of a variant, comprises at least about 10 amino acids, more preferably at least about 20 amino acids, most preferably at least about 30 amino acids, and even more preferably at least about 40 amino acids, in which at least one amino acid is from the tail/head/insertion and at least one amino acid is from the "known protein" portion of a variant. Also optionally, the bridge may comprise any number of amino acids from about 10 to about 40 amino acids (for example, 10, 11, 12, 13...37, 38, 39, 40 amino acids in length, or any number in between).

It should be noted that a bridge cannot be extended beyond the length of the sequence in either direction, and it should be assumed that every bridge description is to be read in such manner that the bridge length does not extend beyond the sequence itself.

Furthermore, bridges are described with regard to a sliding window in certain contexts below. For example, certain descriptions of the bridges feature the following format: a bridge between two edges (in which a portion of the known protein is not present in the variant) may optionally be described as follows: a bridge portion of CONTIG-NAME_P1 (representing the name of the protein), comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise XX (2 amino acids in the center of the bridge, one from each end of the edge), having a structure as follows (numbering according to the sequence of CONTIG-NAME_P1): a sequence starting from any of amino acid numbers 49-x to 49 (for example); and ending at any of amino acid numbers 50 + ((n-2) - x) (for example), in which x varies from 0 to n-2. In this example, it should also be read as including bridges in which n is any number of amino acids between 10-50 amino acids in length. Furthermore, the bridge polypeptide cannot extend beyond the sequence, so it should be read such that 49-x (for example) is not less than 1, nor 50 + ((n-2) – x) (for example) greater than the total sequence length.

In another embodiment, this invention provides antibodies specifically recognizing the splice variants and polypeptide fragments thereof of this invention. Preferably such antibodies differentially recognize splice variants of the present invention but do not recognize a corresponding known protein (such known proteins are discussed with regard to their splice variants in the Examples below).

In another embodiment, this invention provides an isolated nucleic acid molecule encoding for a splice variant according to the present invention, having a nucleotide sequence as set forth in any one of the sequences listed herein, or a sequence complementary thereto. In another embodiment, this invention provides an isolated nucleic acid molecule, having a nucleotide sequence as set forth in any one of the sequences listed herein, or a sequence complementary thereto. In another embodiment, this invention provides an oligonucleotide of at least about 12 nucleotides, specifically hybridizable with the nucleic acid molecules of this invention. In another embodiment, this invention provides vectors, cells, liposomes and compositions comprising the isolated nucleic acids of this invention.

In another embodiment, this invention provides a method for detecting a splice variant according to the present invention in a biological sample, comprising: contacting a biological sample with an antibody specifically recognizing a splice variant according to the present invention under conditions whereby the antibody specifically interacts with the splice variant in the biological sample but do not recognize known corresponding proteins (wherein the known protein is discussed with regard to its splice variant(s) in the Examples below), and detecting said interaction; wherein the presence of an interaction correlates with the presence of a splice variant in the biological sample.

In another embodiment, this invention provides a method for detecting a splice variant nucleic acid sequences in a biological sample, comprising: hybridizing the isolated nucleic acid molecules or oligonucleotide fragments of at least about a minimum length to a nucleic acid material of a biological sample and detecting a hybridization complex; wherein the presence of a hybridization complex correlates with the presence of a splice variant nucleic acid sequence in the biological sample.

According to the present invention, the splice variants described herein are non-limiting examples of markers for diagnosing lung cancer. Each splice variant marker of the present

invention can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, determination of progression, therapy selection and treatment monitoring of lung cancer.

According to optional but preferred embodiments of the present invention, any marker according to the present invention may optionally be used alone or combination. Such a combination may optionally comprise a plurality of markers described herein, optionally including any subcombination of markers, and/or a combination featuring at least one other marker, for example a known marker. Furthermore, such a combination may optionally and preferably be used as described above with regard to determining a ratio between a quantitative or semi-quantitative measurement of any marker described herein to any other marker described herein, and/or any other known marker, and/or any other marker. With regard to such a ratio between any marker described herein (or a combination thereof) and a known marker, more preferably the known marker comprises the "known protein" as described in greater detail below with regard to each cluster or gene.

According to other preferred embodiments of the present invention, a splice variant protein or a fragment thereof, or a splice variant nucleic acid sequence or a fragment thereof, may be featured as a biomarker for detecting lung cancer, such that a biomarker may optionally comprise any of the above.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to a splice variant protein as described herein. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequences of these proteins that are depicted as tails, heads, insertions, edges or bridges. The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to a splice variant of the present invention as described above, optionally for any application.

Non-limiting examples of methods or assays are described below.

The present invention also relates to kits based upon such diagnostic methods or assays.

Nucleic acid sequences and Oligonucleotides

Various embodiments of the present invention encompass nucleic acid sequences described hereinabove; fragments thereof, sequences hybridizable therewith, sequences homologous thereto, sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or artificially induced, either randomly or in a targeted fashion.

The present invention encompasses nucleic acid sequences described herein; fragments thereof, sequences hybridizable therewith, sequences homologous thereto [e.g., at least 50 %, at least 55 %, at least 60%, at least 65 %, at least 70 %, at least 75 %, at least 80 %, at least 85 %, at least 95 % or more say 100 % identical to the nucleic acid sequences set forth below], sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion. The present invention also encompasses homologous nucleic acid sequences (i.e., which form a part of a polynucleotide sequence of the present invention) which include sequence regions unique to the polynucleotides of the present invention.

In cases where the polynucleotide sequences of the present invention encode previously unidentified polypeptides, the present invention also encompasses novel polypeptides or portions thereof, which are encoded by the isolated polynucleotide and respective nucleic acid fragments thereof described hereinabove.

A "nucleic acid fragment" or an "oligonucleotide" or a "polynucleotide" are used herein interchangeably to refer to a polymer of nucleic acids. A polynucleotide sequence of the present invention refers to a single or double stranded nucleic acid sequences which is isolated and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequences (e.g., a combination of the above).

As used herein the phrase "complementary polynucleotide sequence" refers to a sequence, which results from reverse transcription of messenger RNA using a reverse

transcriptase or any other RNA dependent DNA polymerase. Such a sequence can be subsequently amplified *in vivo* or *in vitro* using a DNA dependent DNA polymerase.

As used herein the phrase "genomic polynucleotide sequence" refers to a sequence derived (isolated) from a chromosome and thus it represents a contiguous portion of a chromosome.

As used herein the phrase "composite polynucleotide sequence" refers to a sequence, which is composed of genomic and cDNA sequences. A composite sequence can include some exonal sequences required to encode the polypeptide of the present invention, as well as some intronic sequences interposing therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

Preferred embodiments of the present invention encompass oligonucleotide probes.

An example of an oligonucleotide probe which can be utilized by the present invention is a single stranded polynucleotide which includes a sequence complementary to the unique sequence region of any variant according to the present invention, including but not limited to a nucleotide sequence coding for an amino sequence of a bridge, tail, head and/or insertion according to the present invention, and/or the equivalent portions of any nucleotide sequence given herein (including but not limited to a nucleotide sequence of a node, segment or amplicon described herein).

Alternatively, an oligonucleotide probe of the present invention can be designed to hybridize with a nucleic acid sequence encompassed by any of the above nucleic acid sequences, particularly the portions specified above, including but not limited to a nucleotide sequence coding for an amino sequence of a bridge, tail, head and/or insertion according to the present invention, and/or the equivalent portions of any nucleotide sequence given herein (including but not limited to a nucleotide sequence of a node, segment or amplicon described herein).

Oligonucleotides designed according to the teachings of the present invention can be generated according to any oligonucleotide synthesis method known in the art such as enzymatic synthesis or solid phase synthesis. Equipment and reagents for executing solid-phase synthesis are commercially available from, for example, Applied Biosystems. Any other means for such synthesis may also be employed; the actual synthesis of the oligonucleotides is well within the capabilities of one skilled in the art and can be accomplished via established methodologies as

detailed in, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988) and "Oligonucleotide Synthesis" Gait, M. J., ed. (1984) utilizing solid phase chemistry, e.g. cyanoethyl phosphoramidite followed by deprotection, desalting and purification by for example, an automated trityl-on method or HPLC.

Oligonucleotides used according to this aspect of the present invention are those having a length selected from a range of about 10 to about 200 bases preferably about 15 to about 150 bases, more preferably about 20 to about 100 bases, most preferably about 20 to about 50 bases. Preferably, the oligonucleotide of the present invention features at least 17, at least 18, at least 19, at least 20, at least 22, at least 25, at least 30 or at least 40, bases specifically hybridizable with the biomarkers of the present invention.

The oligonucleotides of the present invention may comprise heterocylic nucleosides consisting of purines and the pyrimidines bases, bonded in a 3' to 5' phosphodiester linkage.

Preferably used oligonucleotides are those modified at one or more of the backbone, internucleoside linkages or bases, as is broadly described hereinunder.

Specific examples of preferred oligonucleotides useful according to this aspect of the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone, as disclosed in U.S. Pat. NOs: 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466, 677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkyl phosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside

units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms can also be used.

Alternatively, modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts, as disclosed in U.S. Pat. Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623, 070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439.

Other oligonucleotides which can be used according to the present invention, are those modified in both sugar and the internucleoside linkage, i.e., the backbone, of the nucleotide units are replaced with novel groups. The base units are maintained for complementation with the appropriate polynucleotide target. An example for such an oligonucleotide mimetic, includes peptide nucleic acid (PNA). United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Other backbone modifications, which can be used in the present invention are disclosed in U.S. Pat. No: 6,303,374.

Oligonucleotides of the present invention may also include base modifications or substitutions. As used herein, "unmodified" or "natural" bases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified bases include but are not limited to other synthetic and natural bases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil

(pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further bases particularly useful for increasing the binding affinity of the oligomeric compounds of the invention include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 °C and are presently preferred base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more moieties or conjugates, which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-S-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety, as disclosed in U.S. Pat. No: 6,303,374.

It is not necessary for all positions in a given oligonucleotide molecule to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide.

It will be appreciated that oligonucleotides of the present invention may include further modifications for more efficient use as diagnostic agents and/or to increase bioavailability, therapeutic efficacy and reduce cytotoxicity.

To enable cellular expression of the polynucleotides of the present invention, a nucleic acid construct according to the present invention may be used, which includes at least a coding region of one of the above nucleic acid sequences, and further includes at least one cis acting regulatory element. As used herein, the phrase "cis acting regulatory element" refers to a polynucleotide sequence, preferably a promoter, which binds a trans acting regulator and regulates the transcription of a coding sequence located downstream thereto.

Any suitable promoter sequence can be used by the nucleic acid construct of the present invention.

Preferably, the promoter utilized by the nucleic acid construct of the present invention is active in the specific cell population transformed. Examples of cell type-specific and/or tissue-specific promoters include promoters such as albumin that is liver specific, lymphoid specific promoters [Calame et al., (1988) Adv. Immunol. 43:235-275]; in particular promoters of T-cell receptors [Winoto et al., (1989) EMBO J. 8:729-733] and immunoglobulins; [Banerji et al. (1983) Cell 33729-740], neuron-specific promoters such as the neurofilament promoter [Byrne et al. (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477], pancreas-specific promoters [Edlunch et al. (1985) Science 230:912-916] or mammary gland-specific promoters such as the milk whey promoter (U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). The nucleic acid construct of the present invention can further include an enhancer, which can be adjacent or distant to the promoter sequence and can function in up regulating the transcription therefrom.

The nucleic acid construct of the present invention preferably further includes an appropriate selectable marker and/or an origin of replication. Preferably, the nucleic acid construct utilized is a shuttle vector, which can propagate both in *E. coli* (wherein the construct comprises an appropriate selectable marker and origin of replication) and be compatible for propagation in cells, or integration in a gene and a tissue of choice. The construct according to the present invention can be, for example, a plasmid, a bacmid, a phagemid, a cosmid, a phage, a virus or an artificial chromosome.

Examples of suitable constructs include, but are not limited to, pcDNA3, pcDNA3.1 (+/-), pGL3, PzeoSV2 (+/-), pDisplay, pEF/myc/cyto, pCMV/myc/cyto each of which is commercially available from Invitrogen Co. (www.invitrogen.com). Examples of retroviral vector and packaging systems are those sold by Clontech, San Diego, Calif., includingRetro-X vectors pLNCX and pLXSN, which permit cloning into multiple cloning sites and the trasgene is transcribed from CMV promoter. Vectors derived from Mo-MuLV are also included such as pBabe, where the transgene will be transcribed from the 5'LTR promoter.

Currently preferred in vivo nucleic acid transfer techniques include transfection with viral or non-viral constructs, such as adenovirus, lentivirus, Herpes simplex I virus, or adeno-associated virus (AAV) and lipid-based systems. Useful lipids for lipid-mediated transfer of the

gene are, for example, DOTMA, DOPE, and DC-Chol [Tonkinson et al., Cancer Investigation, 14(1): 54-65 (1996)]. The most preferred constructs for use in gene therapy are viruses, most preferably adenoviruses, AAV, lentiviruses, or retroviruses. A viral construct such as a retroviral construct includes at least one transcriptional promoter/enhancer or locus-defining element(s), or other elements that control gene expression by other means such as alternate splicing, nuclear RNA export, or post-translational modification of messenger. Such vector constructs also include a packaging signal, long terminal repeats (LTRs) or portions thereof, and positive and negative strand primer binding sites appropriate to the virus used, unless it is already present in the viral construct. In addition, such a construct typically includes a signal sequence for secretion of the peptide from a host cell in which it is placed. Preferably the signal sequence for this purpose is a mammalian signal sequence or the signal sequence of the polypeptide variants of the present invention. Optionally, the construct may also include a signal that directs polyadenylation, as well as one or more restriction sites and a translation termination sequence. By way of example, such constructs will typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second-strand DNA synthesis, and a 3' LTR or a portion thereof. Other vectors can be used that are non-viral, such as cationic lipids, polylysine, and dendrimers.

Hybridization assays

Detection of a nucleic acid of interest in a biological sample may optionally be effected by hybridization-based assays using an oligonucleotide probe (non-limiting examples of probes according to the present invention were previously described).

Traditional hybridization assays include PCR, RT-PCR, Real-time PCR, RNase protection, in-situ hybridization, primer extension, Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection) (NAT type assays are described in greater detail below). More recently, PNAs have been described (Nielsen et al. 1999, Current Opin. Biotechnol. 10:71-75). Other detection methods include kits containing probes on a dipstick setup and the like.

Hybridization based assays which allow the detection of a variant of interest (i.e., DNA or RNA) in a biological sample rely on the use of oligonucleotides which can be 10, 15, 20, or

30 to 100 nucleotides long preferably from 10 to 50, more preferably from 40 to 50 nucleotides long.

Thus, the isolated polynucleotides (oligonucleotides) of the present invention are preferably hybridizable with any of the herein described nucleic acid sequences under moderate to stringent hybridization conditions.

Moderate to stringent hybridization conditions are characterized by a hybridization solution such as containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and 5 x 10^6 cpm ^{32}P labeled probe, at 65 °C, with a final wash solution of 0.2 x SSC and 0.1 % SDS and final wash at 65°C and whereas moderate hybridization is effected using a hybridization solution containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and 5 x 10^6 cpm ^{32}P labeled probe, at 65 °C, with a final wash solution of 1 x SSC and 0.1 % SDS and final wash at 50 °C.

More generally, hybridization of short nucleic acids (below 200 bp in length, e.g. 17-40 bp in length) can be effected using the following exemplary hybridization protocols which can be modified according to the desired stringency; *(i)* hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 µg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 1 - 1.5 °C below the T_m, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the T_m; *(ii)* hybridization solution of 6 x SSC and 0.1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 µg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature of 2 - 2.5 °C below the T_m, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS at 1 - 1.5 °C below the T_m, final wash solution of 6 x SSC, and final wash at 22 °C; *(iii)* hybridization solution of 6 x SSC and 1 % SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5 % SDS, 100 µg/ml denatured salmon sperm DNA and 0.1 % nonfat dried milk, hybridization temperature.

The detection of hybrid duplexes can be carried out by a number of methods. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Such labels refer to radioactive, fluorescent, biological or enzymatic

tags or labels of standard use in the art. A label can be conjugated to either the oligonucleotide probes or the nucleic acids derived from the biological sample.

Probes can be labeled according to numerous well known methods. Non-limiting examples of radioactive labels include 3H, 14C, 32P, and 35S. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents, enzymes, and antibodies. Other detectable markers for use with probes, which can enable an increase in sensitivity of the method of the invention, include biotin and radio-nucleotides. It will become evident to the person of ordinary skill that the choice of a particular label dictates the manner in which it is bound to the probe.

For example, oligonucleotides of the present invention can be labeled subsequent to synthesis, by incorporating biotinylated dNTPs or rNTP, or some similar means (e.g., photo-cross-linking a psoralen derivative of biotin to RNAs), followed by addition of labeled streptavidin (e.g., phycoerythrin-conjugated streptavidin) or the equivalent. Alternatively, when fluorescently-labeled oligonucleotide probes are used, fluorescein, lissamine, phycoerythrin, rhodamine (Perkin Elmer Cetus), Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX (Amersham) and others [e.g., Kricka et al. (1992), Academic Press San Diego, Calif] can be attached to the oligonucleotides.

Those skilled in the art will appreciate that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the oligonucleotide primers and probes.

It will be appreciated that a variety of controls may be usefully employed to improve accuracy of hybridization assays. For instance, samples may be hybridized to an irrelevant probe and treated with RNAse A prior to hybridization, to assess false hybridization.

Although the present invention is not specifically dependent on the use of a label for the detection of a particular nucleic acid sequence, such a label might be beneficial, by increasing the sensitivity of the detection. Furthermore, it enables automation. Probes can be labeled according to numerous well known methods.

As commonly known, radioactive nucleotides can be incorporated into probes of the invention by several methods. Non-limiting examples of radioactive labels include ^{3}H, ^{14}C, ^{32}P, and ^{35}S.

Those skilled in the art will appreciate that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the oligonucleotide primers and probes.

It will be appreciated that a variety of controls may be usefully employed to improve accuracy of hybridization assays.

Probes of the invention can be utilized with naturally occurring sugar-phosphate backbones as well as modified backbones including phosphorothioates, dithionates, alkyl phosphonates and a-nucleotides and the like. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA), and preferably of DNA.

NAT Assays

Detection of a nucleic acid of interest in a biological sample may also optionally be effected by NAT-based assays, which involve nucleic acid amplification technology, such as PCR for example (or variations thereof such as real-time PCR for example).

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to (hybridizing with) a target sequence, thereby creating a double stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14 Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the q3 replicase system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, supra).

The terminology "amplification pair" (or "primer pair") refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes

include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

In one particular embodiment, amplification of a nucleic acid sample from a patient is amplified under conditions which favor the amplification of the most abundant differentially expressed nucleic acid. In one preferred embodiment, RT-PCR is carried out on an mRNA sample from a patient under conditions which favor the amplification of the most abundant mRNA. In another preferred embodiment, the amplification of the differentially expressed nucleic acids is carried out simultaneously. It will be realized by a person skilled in the art that such methods could be adapted for the detection of differentially expressed proteins instead of differentially expressed nucleic acid sequences.

The nucleic acid (i.e. DNA or RNA) for practicing the present invention may be obtained according to well known methods.

Oligonucleotide primers of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted genomes employed. Optionally, the oligonucleotide primers are at least 12 nucleotides in length, preferably between 15 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide primers can be designed by taking into consideration the melting point of hybridization thereof with its targeted sequence (Sambrook et al., 1989, Molecular Cloning -A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

It will be appreciated that antisense oligonucleotides may be employed to quantify expression of a splice isoform of interest. Such detection is effected at the pre-mRNA level. Essentially the ability to quantitate transcription from a splice site of interest can be effected based on splice site accessibility. Oligonucleotides may compete with splicing factors for the splice site sequences. Thus, low activity of the antisense oligonucleotide is indicative of splicing activity.

The polymerase chain reaction and other nucleic acid amplification reactions are well known in the art (various non-limiting examples of these reactions are described in greater detail below). The pair of oligonucleotides according to this aspect of the present invention are

preferably selected to have compatible melting temperatures (Tm), e.g., melting temperatures which differ by less than that 7 °C, preferably less than 5 °C, more preferably less than 4 °C, most preferably less than 3 °C, ideally between 3 °C and 0 °C.

Polymerase Chain Reaction (PCR): The polymerase chain reaction (PCR), as described in U.S. Pat. Nos. 4,683,195 and 4,683,202 to Mullis and Mullis *et al.*, is a method of increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. This technology provides one approach to the problems of low target sequence concentration. PCR can be used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves the introduction of a molar excess of two oligonucleotide primers which are complementary to their respective strands of the double-stranded target sequence to the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of denaturation, hybridization (annealing), and polymerase extension (elongation) can be repeated as often as needed, in order to obtain relatively high concentrations of a segment of the desired target sequence.

The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and, therefore, this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified."

Ligase Chain Reaction (LCR or LAR): The ligase chain reaction [LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR)] has developed into a well-recognized alternative method of amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation, and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes: see for example Segev, PCT Publication No. W09001069 A1 (1990). However,

because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

Self-Sustained Synthetic Reaction (3SR/NASBA): The self-sustained sequence replication reaction (3SR) is a transcription-based in vitro amplification system that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection. In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo-and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (e.g., 200-300 base pairs).

Q-Beta (Qβ) *Replicase:* In this method, a probe which recognizes the sequence of interest is attached to the replicatable RNA template for Qβ replicase. A previously identified major problem with false positives resulting from the replication of unhybridized probes has been addressed through use of a sequence-specific ligation step. However, available thermostable DNA ligases are not effective on this RNA substrate, so the ligation must be performed by T4 DNA ligase at low temperatures (37 degrees C.). This prevents the use of high temperature as a means of achieving specificity as in the LCR, the ligation event can be used to detect a mutation at the junction site, but not elsewhere.

A successful diagnostic method must be very specific. A straight-forward method of controlling the specificity of nucleic acid hybridization is by controlling the temperature of the reaction. While the 3SR/NASBA, and Qβ systems are all able to generate a large quantity of signal, one or more of the enzymes involved in each cannot be used at high temperature (i.e., > 55 degrees C). Therefore the reaction temperatures cannot be raised to prevent non-specific hybridization of the probes. If probes are shortened in order to make them melt more easily at low temperatures, the likelihood of having more than one perfect match in a complex genome increases. For these reasons, PCR and LCR currently dominate the research field in detection technologies.

The basis of the amplification procedure in the PCR and LCR is the fact that the products of one cycle become usable templates in all subsequent cycles, consequently doubling the population with each cycle. The final yield of any such doubling system can be expressed as: $(1+X)^n = y$, where "X" is the mean efficiency (percent copied in each cycle), "n" is the number of cycles, and "y" is the overall efficiency, or yield of the reaction. If every copy of a target DNA is utilized as a template in every cycle of a polymerase chain reaction, then the mean efficiency is 100 %. If 20 cycles of PCR are performed, then the yield will be 2^{20}, or 1,048,576 copies of the starting material. If the reaction conditions reduce the mean efficiency to 85 %, then the yield in those 20 cycles will be only 1.85^{20}, or 220,513 copies of the starting material. In other words, a PCR running at 85 % efficiency will yield only 21 % as much final product, compared to a reaction running at 100 % efficiency. A reaction that is reduced to 50 % mean efficiency will yield less than 1 % of the possible product.

In practice, routine polymerase chain reactions rarely achieve the theoretical maximum yield, and PCRs are usually run for more than 20 cycles to compensate for the lower yield. At 50 % mean efficiency, it would take 34 cycles to achieve the million-fold amplification theoretically possible in 20, and at lower efficiencies, the number of cycles required becomes prohibitive. In addition, any background products that amplify with a better mean efficiency than the intended target will become the dominant products.

Also, many variables can influence the mean efficiency of PCR, including target DNA length and secondary structure, primer length and design, primer and dNTP concentrations, and buffer composition, to name but a few. Contamination of the reaction with exogenous DNA (e.g., DNA spilled onto lab surfaces) or cross-contamination is also a major consideration. Reaction conditions must be carefully optimized for each different primer pair and target sequence, and the process can take days, even for an experienced investigator. The laboriousness of this process, including numerous technical considerations and other factors, presents a significant drawback to using PCR in the clinical setting. Indeed, PCR has yet to penetrate the clinical market in a significant way. The same concerns arise with LCR, as LCR must also be optimized to use different oligonucleotide sequences for each target sequence. In addition, both methods require expensive equipment, capable of precise temperature cycling.

Many applications of nucleic acid detection technologies, such as in studies of allelic variation, involve not only detection of a specific sequence in a complex background, but also

the discrimination between sequences with few, or single, nucleotide differences. One method of the detection of allele-specific variants by PCR is based upon the fact that it is difficult for Taq polymerase to synthesize a DNA strand when there is a mismatch between the template strand and the 3' end of the primer. An allele-specific variant may be detected by the use of a primer that is perfectly matched with only one of the possible alleles; the mismatch to the other allele acts to prevent the extension of the primer, thereby preventing the amplification of that sequence. This method has a substantial limitation in that the base composition of the mismatch influences the ability to prevent extension across the mismatch, and certain mismatches do not prevent extension or have only a minimal effect.

A similar 3'-mismatch strategy is used with greater effect to prevent ligation in the LCR. Any mismatch effectively blocks the action of the thermostable ligase, but LCR still has the drawback of target-independent background ligation products initiating the amplification. Moreover, the combination of PCR with subsequent LCR to identify the nucleotides at individual positions is also a clearly cumbersome proposition for the clinical laboratory.

The direct detection method according to various preferred embodiments of the present invention may be, for example a cycling probe reaction (CPR) or a branched DNA analysis.

When a sufficient amount of a nucleic acid to be detected is available, there are advantages to detecting that sequence directly, instead of making more copies of that target, (e.g., as in PCR and LCR). Most notably, a method that does not amplify the signal exponentially is more amenable to quantitative analysis. Even if the signal is enhanced by attaching multiple dyes to a single oligonucleotide, the correlation between the final signal intensity and amount of target is direct. Such a system has an additional advantage that the products of the reaction will not themselves promote further reaction, so contamination of lab surfaces by the products is not as much of a concern. Recently devised techniques have sought to eliminate the use of radioactivity and/or improve the sensitivity in automatable formats. Two examples are the "Cycling Probe Reaction" (CPR), and "Branched DNA" (bDNA).

Cycling probe reaction (CPR): The cycling probe reaction (CPR), uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the remainder of the probe from the target DNA and allowing another probe

molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may carried through sample preparation.

Branched DNA: Branched DNA (bDNA), involves oligonucleotides with branched structures that allow each individual oligonucleotide to carry 35 to 40 labels (e.g., alkaline phosphatase enzymes). While this enhances the signal from a hybridization event, signal from non-specific binding is similarly increased.

The detection of at least one sequence change according to various preferred embodiments of the present invention may be accomplished by, for example restriction fragment length polymorphism (RFLP analysis), allele specific oligonucleotide (ASO) analysis, Denaturing/Temperature Gradient Gel Electrophoresis (DGGE/TGGE), Single-Strand Conformation Polymorphism (SSCP) analysis or Dideoxy fingerprinting (ddF).

The demand for tests which allow the detection of specific nucleic acid sequences and sequence changes is growing rapidly in clinical diagnostics. As nucleic acid sequence data for genes from humans and pathogenic organisms accumulates, the demand for fast, cost-effective, and easy-to-use tests for as yet mutations within specific sequences is rapidly increasing.

A handful of methods have been devised to scan nucleic acid segments for mutations. One option is to determine the entire gene sequence of each test sample (e.g., a bacterial isolate). For sequences under approximately 600 nucleotides, this may be accomplished using amplified material (e.g., PCR reaction products). This avoids the time and expense associated with cloning the segment of interest. However, specialized equipment and highly trained personnel are required, and the method is too labor-intense and expensive to be practical and effective in the clinical setting.

In view of the difficulties associated with sequencing, a given segment of nucleic acid may be characterized on several other levels. At the lowest resolution, the size of the molecule can be determined by electrophoresis by comparison to a known standard run on the same gel. A more detailed picture of the molecule may be achieved by cleavage with combinations of restriction enzymes prior to electrophoresis, to allow construction of an ordered map. The presence of specific sequences within the fragment can be detected by hybridization of a labeled probe, or the precise nucleotide sequence can be determined by partial chemical degradation or by primer extension in the presence of chain-terminating nucleotide analogs.

Restriction fragment length polymorphism (RFLP): For detection of single-base differences between like sequences, the requirements of the analysis are often at the highest level of resolution. For cases in which the position of the nucleotide in question is known in advance, several methods have been developed for examining single base changes without direct sequencing. For example, if a mutation of interest happens to fall within a restriction recognition sequence, a change in the pattern of digestion can be used as a diagnostic tool (e.g., restriction fragment length polymorphism [RFLP] analysis).

Single point mutations have been also detected by the creation or destruction of RFLPs. Mutations are detected and localized by the presence and size of the RNA fragments generated by cleavage at the mismatches. Single nucleotide mismatches in DNA heteroduplexes are also recognized and cleaved by some chemicals, providing an alternative strategy to detect single base substitutions, generically named the "Mismatch Chemical Cleavage" (MCC). However, this method requires the use of osmium tetroxide and piperidine, two highly noxious chemicals which are not suited for use in a clinical laboratory.

RFLP analysis suffers from low sensitivity and requires a large amount of sample. When RFLP analysis is used for the detection of point mutations, it is, by its nature, limited to the detection of only those single base changes which fall within a restriction sequence of a known restriction endonuclease. Moreover, the majority of the available enzymes have 4 to 6 base-pair recognition sequences, and cleave too frequently for many large-scale DNA manipulations. Thus, it is applicable only in a small fraction of cases, as most mutations do not fall within such sites.

A handful of rare-cutting restriction enzymes with 8 base-pair specificities have been isolated and these are widely used in genetic mapping, but these enzymes are few in number, are limited to the recognition of G+C-rich sequences, and cleave at sites that tend to be highly clustered. Recently, endonucleases encoded by group I introns have been discovered that might have greater than 12 base-pair specificity, but again, these are few in number.

Allele specific oligonucleotide (ASO): If the change is not in a recognition sequence, then allele-specific oligonucleotides (ASOs), can be designed to hybridize in proximity to the mutated nucleotide, such that a primer extension or ligation event can bused as the indicator of a match or a mis-match. Hybridization with radioactively labeled allelic specific oligonucleotides (ASO) also has been applied to the detection of specific point mutations. The method is based

on the differences in the melting temperature of short DNA fragments differing by a single nucleotide. Stringent hybridization and washing conditions can differentiate between mutant and wild-type alleles. The ASO approach applied to PCR products also has been extensively utilized by various researchers to detect and characterize point mutations in ras genes and gsp/gip oncogenes. Because of the presence of various nucleotide changes in multiple positions, the ASO method requires the use of many oligonucleotides to cover all possible oncogenic mutations.

With either of the techniques described above (i.e., RFLP and ASO), the precise location of the suspected mutation must be known in advance of the test. That is to say, they are inapplicable when one needs to detect the presence of a mutation within a gene or sequence of interest.

Denaturing/Temperature Gradient Gel Electrophoresis (DGGE/TGGE): Two other methods rely on detecting changes in electrophoretic mobility in response to minor sequence changes. One of these methods, termed "Denaturing Gradient Gel Electrophoresis" (DGGE) is based on the observation that slightly different sequences will display different patterns of local melting when electrophoretically resolved on a gradient gel. In this manner, variants can be distinguished, as differences in melting properties of homoduplexes versus heteroduplexes differing in a single nucleotide can detect the presence of mutations in the target sequences because of the corresponding changes in their electrophoretic mobilities. The fragments to be analyzed, usually PCR products, are "clamped" at one end by a long stretch of G-C base pairs (30-80) to allow complete denaturation of the sequence of interest without complete dissociation of the strands. The attachment of a GC "clamp" to the DNA fragments increases the fraction of mutations that can be recognized by DGGE. Attaching a GC clamp to one primer is critical to ensure that the amplified sequence has a low dissociation temperature. Modifications of the technique have been developed, using temperature gradients, and the method can be also applied to RNA:RNA duplexes.

Limitations on the utility of DGGE include the requirement that the denaturing conditions must be optimized for each type of DNA to be tested. Furthermore, the method requires specialized equipment to prepare the gels and maintain the needed high temperatures during electrophoresis. The expense associated with the synthesis of the clamping tail on one oligonucleotide for each sequence to be tested is also a major consideration. In addition, long

running times are required for DGGE. The long running time of DGGE was shortened in a modification of DGGE called constant denaturant gel electrophoresis (CDGE). CDGE requires that gels be performed under different denaturant conditions in order to reach high efficiency for the detection of mutations.

A technique analogous to DGGE, termed temperature gradient gel electrophoresis (TGGE), uses a thermal gradient rather than a chemical denaturant gradient. TGGE requires the use of specialized equipment which can generate a temperature gradient perpendicularly oriented relative to the electrical field. TGGE can detect mutations in relatively small fragments of DNA therefore scanning of large gene segments requires the use of multiple PCR products prior to running the gel.

Single-Strand Conformation Polymorphism (SSCP): Another common method, called "Single-Strand Conformation Polymorphism" (SSCP) was developed by Hayashi, Sekya and colleagues and is based on the observation that single strands of nucleic acid can take on characteristic conformations in non-denaturing conditions, and these conformations influence electrophoretic mobility. The complementary strands assume sufficiently different structures that one strand may be resolved from the other. Changes in sequences within the fragment will also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations.

The SSCP process involves denaturing a DNA segment (e.g., a PCR product) that is labeled on both strands, followed by slow electrophoretic separation on a non-denaturing polyacrylamide gel, so that intra-molecular interactions can form and not be disturbed during the run. This technique is extremely sensitive to variations in gel composition and temperature. A serious limitation of this method is the relative difficulty encountered in comparing data generated in different laboratories, under apparently similar conditions.

Dideoxy fingerprinting (ddF): The dideoxy fingerprinting (ddF) is another technique developed to scan genes for the presence of mutations. The ddF technique combines components of Sanger dideoxy sequencing with SSCP. A dideoxy sequencing reaction is performed using one dideoxy terminator and then the reaction products are electrophoresed on nondenaturing polyacrylamide gels to detect alterations in mobility of the termination segments as in SSCP analysis. While ddF is an improvement over SSCP in terms of increased sensitivity, ddF requires the use of expensive dideoxynucleotides and this technique is still limited to the

analysis of fragments of the size suitable for SSCP (i.e., fragments of 200-300 bases for optimal detection of mutations).

In addition to the above limitations, all of these methods are limited as to the size of the nucleic acid fragment that can be analyzed. For the direct sequencing approach, sequences of greater than 600 base pairs require cloning, with the consequent delays and expense of either deletion sub-cloning or primer walking, in order to cover the entire fragment. SSCP and DGGE have even more severe size limitations. Because of reduced sensitivity to sequence changes, these methods are not considered suitable for larger fragments. Although SSCP is reportedly able to detect 90 % of single-base substitutions within a 200 base-pair fragment, the detection drops to less than 50 % for 400 base pair fragments. Similarly, the sensitivity of DGGE decreases as the length of the fragment reaches 500 base-pairs. The ddF technique, as a combination of direct sequencing and SSCP, is also limited by the relatively small size of the DNA that can be screened.

According to a presently preferred embodiment of the present invention the step of searching for any of the nucleic acid sequences described here, in tumor cells or in cells derived from a cancer patient is effected by any suitable technique, including, but not limited to, nucleic acid sequencing, polymerase chain reaction, ligase chain reaction, self-sustained synthetic reaction, Qβ-Replicase, cycling probe reaction, branched DNA, restriction fragment length polymorphism analysis, mismatch chemical cleavage, heteroduplex analysis, allele-specific oligonucleotides, denaturing gradient gel electrophoresis, constant denaturant gel electrophoresis, temperature gradient gel electrophoresis and dideoxy fingerprinting.

Detection may also optionally be performed with a chip or other such device. The nucleic acid sample which includes the candidate region to be analyzed is preferably isolated, amplified and labeled with a reporter group. This reporter group can be a fluorescent group such as phycoerythrin. The labeled nucleic acid is then incubated with the probes immobilized on the chip using a fluidics station. describe the fabrication of fluidics devices and particularly microcapillary devices, in silicon and glass substrates.

Once the reaction is completed, the chip is inserted into a scanner and patterns of hybridization are detected. The hybridization data is collected, as a signal emitted from the reporter groups already incorporated into the nucleic acid, which is now bound to the probes

attached to the chip. Since the sequence and position of each probe immobilized on the chip is known, the identity of the nucleic acid hybridized to a given probe can be determined.

It will be appreciated that when utilized along with automated equipment, the above described detection methods can be used to screen multiple samples for a disease and/or pathological condition both rapidly and easily.

Amino acid sequences and peptides

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, e.g., by the addition of carbohydrate residues to form glycoproteins. The terms "polypeptide," "peptide" and "protein" include glycoproteins, as well as non-glycoproteins.

Polypeptide products can be biochemically synthesized such as by employing standard solid phase techniques. Such methods include but are not limited to exclusive solid phase synthesis, partial solid phase synthesis methods, fragment condensation, classical solution synthesis. These methods are preferably used when the peptide is relatively short (i.e., 10 kDa) and/or when it cannot be produced by recombinant techniques (i.e., not encoded by a nucleic acid sequence) and therefore involves different chemistry.

Solid phase polypeptide synthesis procedures are well known in the art and further described by John Morrow Stewart and Janis Dillaha Young, Solid Phase Peptide Syntheses (2nd Ed., Pierce Chemical Company, 1984).

Synthetic polypeptides can optionally be purified by preparative high performance liquid chromatography [Creighton T. (1983) Proteins, structures and molecular principles. WH Freeman and Co. N.Y.], after which their composition can be confirmed via amino acid sequencing.

In cases where large amounts of a polypeptide are desired, it can be generated using recombinant techniques such as described by Bitter et al., (1987) Methods in Enzymol. 153:516-544, Studier et al. (1990) Methods in Enzymol. 185:60-89, Brisson et al. (1984) Nature 310:511-514, Takamatsu et al. (1987) EMBO J. 6:307-311, Coruzzi et al. (1984) EMBO J. 3:1671-1680 and Brogli et al., (1984) Science 224:838-843, Gurley et al. (1986) Mol. Cell. Biol. 6:559-565

and Weissbach & Weissbach, 1988, Methods for Plant Molecular Biology, Academic Press, NY, Section VIII, pp 421-463.

The present invention also encompasses polypeptides encoded by the polynucleotide sequences of the present invention, as well as polypeptides according to the amino acid sequences described herein. The present invention also encompasses homologues of these polypeptides, such homologues can be at least 50 %, at least 55 %, at least 60%, at least 65 %, at least 70 %, at least 75 %, at least 80 %, at least 85 %, at least 95 % or more say 100 % homologous to the amino acid sequences set forth below, as can be determined using BlastP software of the National Center of Biotechnology Information (NCBI) using default parameters, optionally and preferably including the following: filtering on (this option filters repetitive or low-complexity sequences from the query using the Seg (protein) program), scoring matrix is BLOSUM62 for proteins, word size is 3, E value is 10, gap costs are 11, 1 (initialization and extension), and number of alignments shown is 50. Optionally, nucleic acid sequence identity/homology may be determined by using BlastN software of the National Center of Biotechnology Information (NCBI) using default parameters, which preferably include using the DUST filter program, and also preferably include having an E value of 10, filtering low complexity sequences and a word size of 11. Finally, the present invention also encompasses fragments of the above described polypeptides and polypeptides having mutations, such as deletions, insertions or substitutions of one or more amino acids, either naturally occurring or artificially induced, either randomly or in a targeted fashion.

It will be appreciated that peptides identified according the present invention may be degradation products, synthetic peptides or recombinant peptides as well as peptidomimetics, typically, synthetic peptides and peptoids and semipeptoids which are peptide analogs, which may have, for example, modifications rendering the peptides more stable while in a body or more capable of penetrating into cells. Such modifications include, but are not limited to N terminus modification, C terminus modification, peptide bond modification, including, but not limited to, CH2-NH, CH2-S, CH2-S=O, O=C-NH, CH2-O, CH2-CH2, S=C-NH, CH=CH or CF=CH, backbone modifications, and residue modification. Methods for preparing peptidomimetic compounds are well known in the art and are specified. Further details in this respect are provided hereinunder.

Peptide bonds (-CO-NH-) within the peptide may be substituted, for example, by N-methylated bonds (-N(CH3)-CO-), ester bonds (-C(R)H-C-O-O-C(R)-N-), ketomethylen bonds (-CO-CH2-), α-aza bonds (-NH-N(R)-CO-), wherein R is any alkyl, e.g., methyl, carba bonds (-CH2-NH-), hydroxyethylene bonds (-CH(OH)-CH2-), thioamide bonds (-CS-NH-), olefinic double bonds (-CH=CH-), retro amide bonds (-NH-CO-), peptide derivatives (-N(R)-CH2-CO-), wherein R is the "normal" side chain, naturally presented on the carbon atom.

These modifications can occur at any of the bonds along the peptide chain and even at several (2-3) at the same time.

Natural aromatic amino acids, Trp, Tyr and Phe, may be substituted for synthetic non-natural acid such as Phenylglycine, TIC, naphthylelanine (Nol), ring-methylated derivatives of Phe, halogenated derivatives of Phe or o-methyl-Tyr.

In addition to the above, the peptides of the present invention may also include one or more modified amino acids or one or more non-amino acid monomers (e.g. fatty acids, complex carbohydrates etc).

As used herein in the specification and in the claims section below the term "amino acid" or "amino acids" is understood to include the 20 naturally occurring amino acids; those amino acids often modified post-translationally *in vivo*, including, for example, hydroxyproline, phosphoserine and phosphothreonine; and other unusual amino acids including, but not limited to, 2-aminoadipic acid, hydroxylysine, isodesmosine, nor-valine, nor-leucine and ornithine. Furthermore, the term "amino acid" includes both D- and L-amino acids.

Table 1 non-conventional or modified amino acids which can be used with the present invention.

Table 1

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-	Cpro	L-N-methylasparagine	Nmasn

Carboxylate		L-N-methylaspartic acid	Nmasp
Aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-	Norb	L-N-methylglutamine	Nmgin
Carboxylate		L-N-methylglutamic acid	Nmglu
Cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
Cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle
D-tryptophan	Dtrp	L-norvaline	Nva
D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
D-α-methylarginine	Dmarg	α-methylcyclopentylalanine	Mcpen
D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
D-α-methylaspartate	Dmasp	α- methylpenicillamine	Mpen

D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
D-α-methylisoleucine	Dmile	N- amino-α-methylbutyrate	Nmaabu
D-α-methylleucine	Dmleu	α-napthylalanine	Anap
D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
D-α-methylvaline	Dmval	N-cyclododeclglycine	Ncdod
D-α-methylalnine	Dnmala	N-cyclooctylglycine	Ncoct
D-α-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
D-α-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
D-α-methylasparatate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
D-α-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
D-N-methylleucine	Dnmleu	N-(3-indolylyethyl) glycine	Nhtrp
D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser

N-(2-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nva
D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L-α-methylalanine	Mala
L-α-methylarginine	Marg	L-α-methylasparagine	Masn
L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
L-α-methylhistidine	Mhis	L-α-methylhomo phenylalanine	Mhphe
L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
D-N-methylhistidine	Dnmhis	N-(hydroxyethyl)glycine	Nser
D-N-methylisoleucine	Dnmile	N-(imidazolylethyl)glycine	Nhis
D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser

N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L-α-methylalanine	Mala
L-α-methylarginine	Marg	L-α-methylasparagine	Masn
L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
L-α-methylserine	mser	L-α-methylthreonine	Mthr
L-α-methylvaline	Mtrp	L-α-methyltyrosine	Mtyr
L-α-methylleucine	Mval Nnbhm	L-N-methylhomophenylalanine	Nmhphe
N-(N-(2,2-diphenylethyl)		N-(N-(3,3-diphenylpropyl)	
carbamylmethyl-glycine	Nnbhm	carbamylmethyl(1)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl ethylamino)cyclopropane	Nmbc		

Table 1 Cont.

Since the peptides of the present invention are preferably utilized in diagnostics which require the peptides to be in soluble form, the peptides of the present invention preferably include one or more non-natural or natural polar amino acids, including but not limited to serine and threonine which are capable of increasing peptide solubility due to their hydroxyl-containing side chain.

The peptides of the present invention are preferably utilized in a linear form, although it will be appreciated that in cases where cyclicization does not severely interfere with peptide characteristics, cyclic forms of the peptide can also be utilized.

The peptides of present invention can be biochemically synthesized such as by using standard solid phase techniques. These methods include exclusive solid phase synthesis well known in the art, partial solid phase synthesis methods, fragment condensation, classical solution synthesis. These methods are preferably used when the peptide is relatively short (i.e., 10 kDa) and/or when it cannot be produced by recombinant techniques (i.e., not encoded by a nucleic acid sequence) and therefore involves different chemistry.

Synthetic peptides can be purified by preparative high performance liquid chromatography and the composition of which can be confirmed via amino acid sequencing.

In cases where large amounts of the peptides of the present invention are desired, the peptides of the present invention can be generated using recombinant techniques such as described by Bitter et al., (1987) Methods in Enzymol. 153:516-544, Studier et al. (1990) Methods in Enzymol. 185:60-89, Brisson et al. (1984) Nature 310:511-514, Takamatsu et al. (1987) EMBO J. 6:307-311, Coruzzi et al. (1984) EMBO J. 3:1671-1680 and Brogli et al., (1984) Science 224:838-843, Gurley et al. (1986) Mol. Cell. Biol. 6:559-565 and Weissbach & Weissbach, 1988, Methods for Plant Molecular Biology, Academic Press, NY, Section VIII, pp 421-463 and also as described above.

Antibodies

"Antibody" refers to a polypeptide ligand that is preferably substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an epitope (e.g., an antigen). The recognized immunoglobulin genes include the kappa and lambda light chain constant region genes, the alpha, gamma, delta, epsilon and mu

heavy chain constant region genes, and the myriad-immunoglobulin variable region genes. Antibodies exist, e.g., as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. This includes, e.g., Fab' and $F(ab)'_2$ fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies. It also includes polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, or single chain antibodies. "Fc" portion of an antibody refers to that portion of an immunoglobulin heavy chain that comprises one or more heavy chain constant region domains, CH1, CH2 and CH3, but does not include the heavy chain variable region.

The functional fragments of antibodies, such as Fab, F(ab')2, and Fv that are capable of binding to macrophages, are described as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule, can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule that can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab')2 is a dimer of two Fab' fragments held together by two disulfide bonds; (4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (5) Single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain and the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of producing polyclonal and monoclonal antibodies as well as fragments thereof are well known in the art (See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1988, incorporated herein by reference).

Antibody fragments according to the present invention can be prepared by proteolytic hydrolysis of the antibody or by expression in E. coli or mammalian cells (e.g. Chinese hamster ovary cell culture or other protein expression systems) of DNA encoding the fragment. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by

conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted F(ab')2. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulfhydryl groups resulting from cleavage of disulfide linkages, to produce 3.5S Fab' monovalent fragments. Alternatively, an enzymatic cleavage using pepsin produces two monovalent Fab' fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. Pat. Nos. 4,036,945 and 4,331,647, and references contained therein, which patents are hereby incorporated by reference in their entirety. See also Porter, R. R. [Biochem. J. 73: 119-126 (1959)]. Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

Fv fragments comprise an association of VH and VL chains. This association may be noncovalent, as described in Inbar et al. [Proc. Nat'l Acad. Sci. USA 69:2659-62 (19720]. Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde. Preferably, the Fv fragments comprise VH and VL chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the VH and VL domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced into a host cell such as E. coli. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing sFvs are described, for example, by [Whitlow and Filpula, Methods 2: 97-105 (1991); Bird et al., Science 242:423-426 (1988); Pack et al., Bio/Technology 11:1271-77 (1993); and U.S. Pat. No. 4,946,778, which is hereby incorporated by reference in its entirety.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells. See, for example, Larrick and Fry [Methods, 2: 106-10 (1991)].

Humanized forms of non-human (e.g., murine) antibodies are chimeric molecules of immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab') or

other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)].

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as import residues, which are typically taken from an import variable domain. Humanization can be essentially performed following the method of Winter and co-workers [Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)], by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such humanized antibodies are chimeric antibodies (U.S. Pat. No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries [Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)]. The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal

Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)]. Similarly, human antibodies can be made by introduction of human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Pat. Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in the following scientific publications: Marks et al., Bio/Technology 10,: 779-783 (1992); Lonberg et al., Nature 368: 856-859 (1994); Morrison, Nature 368 812-13 (1994); Fishwild et al., Nature Biotechnology 14, 845-51 (1996); Neuberger, Nature Biotechnology 14: 826 (1996); and Lonberg and Huszar, Intern. Rev. Immunol. 13, 65-93 (1995).

Preferably, the antibody of this aspect of the present invention specifically binds at least one epitope of the polypeptide variants of the present invention. As used herein, the term "epitope" refers to any antigenic determinant on an antigen to which the paratope of an antibody binds.

Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or carbohydrate side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics.

Optionally, a unique epitope may be created in a variant due to a change in one or more post-translational modifications, including but not limited to glycosylation and/or phosphorylation, as described below. Such a change may also cause a new epitope to be created, for example through removal of glycosylation at a particular site.

An epitope according to the present invention may also optionally comprise part or all of a unique sequence portion of a variant according to the present invention in combination with at least one other portion of the variant which is not contiguous to the unique sequence portion in the linear polypeptide itself, yet which are able to form an epitope in combination. One or more unique sequence portions may optionally combine with one or more other non-contiguous portions of the variant (including a portion which may have high homology to a portion of the known protein) to form an epitope.

Immunoassays

In another embodiment of the present invention, an immunoassay can be used to qualitatively or quantitatively detect and analyze markers in a sample. This method comprises: providing an antibody that specifically binds to a marker; contacting a sample with the antibody; and detecting the presence of a complex of the antibody bound to the marker in the sample.

To prepare an antibody that specifically binds to a marker, purified protein markers can be used. Antibodies that specifically bind to a protein marker can be prepared using any suitable methods known in the art.

After the antibody is provided, a marker can be detected and/or quantified using any of a number of well recognized immunological binding assays. Useful assays include, for example, an enzyme immune assay (EIA) such as enzyme-linked immunosorbent assay (ELISA), a radioimmune assay (RIA), a Western blot assay, or a slot blot assay see, e.g., U.S. Pat. Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). Generally, a sample obtained from a subject can be contacted with the antibody that specifically binds the marker.

Optionally, the antibody can be fixed to a solid support to facilitate washing and subsequent isolation of the complex, prior to contacting the antibody with a sample. Examples of solid supports include but are not limited to glass or plastic in the form of, e.g., a microtiter plate, a stick, a bead, or a microbead. Antibodies can also be attached to a solid support.

After incubating the sample with antibodies, the mixture is washed and the antibody-marker complex formed can be detected. This can be accomplished by incubating the washed mixture with a detection reagent. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound marker-specific antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the marker are incubated simultaneously with the mixture.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, marker, volume of solution, concentrations and the like. Usually the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10 °C to 40 °C.

The immunoassay can be used to determine a test amount of a marker in a sample from a subject. First, a test amount of a marker in a sample can be detected using the immunoassay methods described above. If a marker is present in the sample, it will form an antibody-marker complex with an antibody that specifically binds the marker under suitable incubation conditions described above. The amount of an antibody-marker complex can optionally be determined by comparing to a standard. As noted above, the test amount of marker need not be measured in absolute units, as long as the unit of measurement can be compared to a control amount and/or signal.

Preferably used are antibodies which specifically interact with the polypeptides of the present invention and not with wild type proteins or other isoforms thereof, for example. Such antibodies are directed, for example, to the unique sequence portions of the polypeptide variants of the present invention, including but not limited to bridges, heads, tails and insertions described in greater detail below. Preferred embodiments of antibodies according to the present invention are described in greater detail with regard to the section entitled "Antibodies".

Radio-immunoassay (RIA): In one version, this method involves precipitation of the desired substrate and in the methods detailed hereinbelow, with a specific antibody and radiolabelled antibody binding protein (e.g., protein A labeled with I^{125}) immobilized on a precipitable carrier such as agarose beads. The number of counts in the precipitated pellet is proportional to the amount of substrate.

In an alternate version of the RIA, a labeled substrate and an unlabelled antibody binding protein are employed. A sample containing an unknown amount of substrate is added in varying amounts. The decrease in precipitated counts from the labeled substrate is proportional to the amount of substrate in the added sample.

Enzyme linked immunosorbent assay (ELISA): This method involves fixation of a sample (e.g., fixed cells or a proteinaceous solution) containing a protein substrate to a surface such as a well of a microtiter plate. A substrate specific antibody coupled to an enzyme is applied and allowed to bind to the substrate. Presence of the antibody is then detected and quantitated by a colorimetric reaction employing the enzyme coupled to the antibody. Enzymes commonly employed in this method include horseradish peroxidase and alkaline phosphatase. If well calibrated and within the linear range of response, the amount of substrate present in the sample

is proportional to the amount of color produced. A substrate standard is generally employed to improve quantitative accuracy.

Western blot: This method involves separation of a substrate from other protein by means of an acrylamide gel followed by transfer of the substrate to a membrane (e.g., nylon or PVDF). Presence of the substrate is then detected by antibodies specific to the substrate, which are in turn detected by antibody binding reagents. Antibody binding reagents may be, for example, protein A, or other antibodies. Antibody binding reagents may be radiolabelled or enzyme linked as described hereinabove. Detection may be by autoradiography, colorimetric reaction or chemiluminescence. This method allows both quantitation of an amount of substrate and determination of its identity by a relative position on the membrane which is indicative of a migration distance in the acrylamide gel during electrophoresis.

Immunohistochemical analysis: This method involves detection of a substrate *in situ* in fixed cells by substrate specific antibodies. The substrate specific antibodies may be enzyme linked or linked to fluorophores. Detection is by microscopy and subjective evaluation. If enzyme linked antibodies are employed, a colorimetric reaction may be required.

Fluorescence activated cell sorting (FACS): This method involves detection of a substrate *in situ* in cells by substrate specific antibodies. The substrate specific antibodies are linked to fluorophores. Detection is by means of a cell sorting machine which reads the wavelength of light emitted from each cell as it passes through a light beam. This method may employ two or more antibodies simultaneously.

Radio-imaging Methods

These methods include but are not limited to, positron emission tomography (PET) single photon emission computed tomography (SPECT). Both of these techniques are non-invasive, and can be used to detect and/or measure a wide variety of tissue events and/or functions, such as detecting cancerous cells for example. Unlike PET, SPECT can optionally be used with two labels simultaneously. SPECT has some other advantages as well, for example with regard to cost and the types of labels that can be used. For example, US Patent No. 6,696,686 describes the use of SPECT for detection of breast cancer, and is hereby incorporated by reference as if fully set forth herein.

Display Libraries

According to still another aspect of the present invention there is provided a display library comprising a plurality of display vehicles (such as phages, viruses or bacteria) each displaying at least 6, at least 7, at least 8, at least 9, at least 10, 10-15, 12-17, 15-20, 15-30 or 20-50 consecutive amino acids derived from the polypeptide sequences of the present invention.

Methods of constructing such display libraries are well known in the art. Such methods are described in, for example, Young AC, *et al.*, "The three-dimensional structures of a polysaccharide binding antibody to Cryptococcus neoformans and its complex with a peptide from a phage display library: implications for the identification of peptide mimotopes" J Mol Biol 1997 Dec 12;274(4):622-34; Giebel LB *et al.* "Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities" Biochemistry 1995 Nov 28;34(47):15430-5; Davies EL *et al.*, "Selection of specific phage-display antibodies using libraries derived from chicken immunoglobulin genes" J Immunol Methods 1995 Oct 12;186(1):125-35; Jones C RT al. "Current trends in molecular recognition and bioseparation" J Chromatogr A 1995 Jul 14;707(1):3-22; Deng SJ *et al.* "Basis for selection of improved carbohydrate-binding single-chain antibodies from synthetic gene libraries" Proc Natl Acad Sci U S A 1995 May 23;92(11):4992-6; and Deng SJ *et al.* "Selection of antibody single-chain variable fragments with improved carbohydrate binding by phage display" J Biol Chem 1994 Apr 1;269(13):9533-8, which are incorporated herein by reference.

The following sections relate to Candidate Marker Examples (first section) and to Experimental Data for these Marker Examples (second section).

CANDIDATE MARKER EXAMPLES SECTION

This Section relates to Examples of sequences according to the present invention, including illustrative methods of selection thereof.

Description of the methodology undertaken to uncover the biomolecular sequences of the present invention

Human ESTs and cDNAs were obtained from GenBank versions 136 (June 15, 2003 ftp.ncbi.nih.gov/genbank/release.notes/gb136.release.notes); NCBI genome assembly of April 2003; RefSeq sequences from June 2003; Genbank version 139 (December 2003); Human

Genome from NCBI (Build 34) (from Oct 2003); and RefSeq sequences from December 2003; and from the LifeSeq library of Incyte Corporation (ESTs only; Wilmington, DE, USA). With regard to GenBank sequences, the human EST sequences from the EST (GBEST) section and the human mRNA sequences from the primate (GBPRI) section were used; also the human nucleotide RefSeq mRNA sequences were used (see for example www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html and for a reference to the EST section, see www.ncbi.nlm.nih.gov/dbEST/; a general reference to dbEST, the EST database in GenBank, may be found in Boguski et al, Nat Genet. 1993 Aug;4(4):332-3; all of which are hereby incorporated by reference as if fully set forth herein).

Novel splice variants were predicted using the LEADS clustering and assembly system as described in Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. Genome Res 12, 1060-7 (2002); US patent No: 6,625,545; and U.S. Pat. Appl. No. 10/426,002, published as US20040101876 on May 27 2004; all of which are hereby incorporated by reference as if fully set forth herein. Briefly, the software cleans the expressed sequences from repeats, vectors and immunoglobulins. It then aligns the expressed sequences to the genome taking alternatively splicing into account and clusters overlapping expressed sequences into "clusters" that represent genes or partial genes.

These were annotated using the GeneCarta (Compugen, Tel-Aviv, Israel) platform. The GeneCarta platform includes a rich pool of annotations, sequence information (particularly of spliced sequences), chromosomal information, alignments, and additional information such as SNPs, gene ontology terms, expression profiles, functional analyses, detailed domain structures, known and predicted proteins and detailed homology reports.

A brief explanation is provided with regard to the method of selecting the candidates. However, it should noted that this explanation is provided for descriptive purposes only, and is not intended to be limiting in any way. The potential markers were identified by a computational process that was designed to find genes and/or their splice variants that are over-expressed in tumor tissues, by using databases of expressed sequences. Various parameters related to the information in the EST libraries, determined according to a manual classification process, were used to assist in locating genes and/or splice variants thereof that are over-expressed in cancerous tissues. The detailed description of the selection method is presented in Example 1

below. The cancer biomarkers selection engine and the following wet validation stages are schematically summarized in Figure 1.

EXAMPLE 1

Identification of differentially expressed gene products – Algorithm

In order to distinguish between differentially expressed gene products and constitutively expressed genes (i.e., house keeping genes) an algorithm based on an analysis of frequencies was configured. A specific algorithm for identification of transcripts over expressed in cancer is described hereinbelow.

Dry analysis

Library annotation – EST libraries are manually classified according to:

- Tissue origin
- Biological source – Examples of frequently used biological sources for construction of EST libraries include cancer cell-lines; normal tissues; cancer tissues; fetal tissues; and others such as normal cell lines and pools of normal cell-lines, cancer cell-lines and combinations thereof. A specific description of abbreviations used below with regard to these tissues/cell lines etc is given above.
- Protocol of library construction – various methods are known in the art for library construction including normalized library construction; non-normalized library construction; subtracted libraries; ORESTES and others. It will be appreciated that at times the protocol of library construction is not indicated.

The following rules are followed:

EST libraries originating from identical biological samples are considered as a single library.

EST libraries which included above-average levels of contamination, such as DNA contamination for example, were eliminated. The presence of such contamination was determined as follows. For each library, the number of unspliced ESTs that are not fully contained within other spliced sequences was counted. If the percentage of such sequences (as compared to all other sequences) was at least 4 standard deviations above the average for all libraries being

analyzed, this library was tagged as being contaminated and was eliminated from further consideration in the below analysis (see also Sorek, R. & Safer, H.M. A novel algorithm for computational identification of contaminated EST libraries. Nucleic Acids Res 31, 1067-74 (2003)for further details).

Clusters (genes) having at least five sequences including at least two sequences from the tissue of interest were analyzed. Splice variants were identified by using the LEADS software package as described above.

EXAMPLE 2

Identification of genes over expressed in cancer.

Two different scoring algorithms were developed.

Libraries score –candidate sequences which are supported by a number of cancer libraries, are more likely to serve as specific and effective diagnostic markers.

The basic algorithm - for each cluster the number of cancer and normal libraries contributing sequences to the cluster was counted. Fisher exact test was used to check if cancer libraries are significantly over-represented in the cluster as compared to the total number of cancer and normal libraries.

Library counting: Small libraries (e.g., less than 1000 sequences) were excluded from consideration unless they participate in the cluster. For this reason, the total number of libraries is actually adjusted for each cluster.

Clones no. score – Generally, when the number of ESTs is much higher in the cancer libraries relative to the normal libraries it might indicate actual over-expression.

The algorithm –

Clone counting: For counting EST clones each library protocol class was given a weight based on our belief of how much the protocol reflects actual expression levels:

(i) non-normalized : 1

(ii) normalized : 0.2

(iii) all other classes : 0.1

Clones number score - The total weighted number of EST clones from cancer libraries was compared to the EST clones from normal libraries. To avoid cases where one library

contributes to the majority of the score, the contribution of the library that gives most clones for a given cluster was limited to 2 clones.

The score was computed as

$$\frac{\frac{c+1}{C}}{\frac{n+1}{N}}$$

where:

c – weighted number of "cancer" clones in the cluster.

C- weighted number of clones in all "cancer" libraries.

n - weighted number of "normal" clones in the cluster.

N- weighted number of clones in all "normal" libraries.

Clones number score significance - Fisher exact test was used to check if EST clones from cancer libraries are significantly over-represented in the cluster as compared to the total number of EST clones from cancer and normal libraries.

Two search approaches were used to find either general cancer-specific candidates or tumor specific candidates.

- Libraries/sequences originating from tumor tissues are counted as well as libraries originating from cancer cell-lines ("normal" cell-lines were ignored).
- Only libraries/sequences originating from tumor tissues are counted

EXAMPLE 3

Identification of tissue specific genes

For detection of tissue specific clusters, tissue libraries/sequences were compared to the total number of libraries/sequences in cluster. Similar statistical tools to those described in above were employed to identify tissue specific genes. Tissue abbreviations are the same as for cancerous tissues, but are indicated with the header "normal tissue".

The algorithm - for each tested tissue T and for each tested cluster the following were examined:

1. Each cluster includes at least 2 libraries from the tissue T. At least 3 clones (weighed - as described above) from tissue T in the cluster; and

2. Clones from the tissue T are at least 40 % from all the clones participating in the tested cluster

Fisher exact test P-values were computed both for library and weighted clone counts to check that the counts are statistically significant.

EXAMPLE 4

Identification of splice variants over expressed in cancer of clusters which are not over expressed in cancer

Cancer-specific splice variants containing a unique region were identified.

Identification of unique sequence regions in splice variants

A Region is defined as a group of adjacent exons that always appear or do not appear together in each splice variant.

A "segment" (sometimes referred also as "seg" or "node") is defined as the shortest contiguous transcribed region without known splicing inside.

Only reliable ESTs were considered for region and segment analysis. An EST was defined as unreliable if:

(i) Unspliced;

(ii) Not covered by RNA;

(iii) Not covered by spliced ESTs; and

(iv) Alignment to the genome ends in proximity of long poly-A stretch or starts in proximity of long poly-T stretch.

Only reliable regions were selected for further scoring. Unique sequence regions were considered reliable if:

(i) Aligned to the genome; and

(ii) Regions supported by more than 2 ESTs.

The algorithm

Each unique sequence region divides the set of transcripts into 2 groups:

(i) Transcripts containing this region (group TA).

(ii) Transcripts not containing this region (group TB).

The set of EST clones of every cluster is divided into 3 groups:

(i) Supporting (originating from) transcripts of group TA (S1).

(ii) Supporting transcripts of group TB (S2).

(iii) Supporting transcripts from both groups (S3).

Library and clones number scores described above were given to S1 group.

Fisher Exact Test P-values were used to check if:

S1 is significantly enriched by cancer EST clones compared to S2; and

S1 is significantly enriched by cancer EST clones compared to cluster background (S1+S2+S3).

Identification of unique sequence regions and division of the group of transcripts accordingly is illustrated in Figure 2. Each of these unique sequence regions corresponds to a segment, also termed herein a "node".

Region 1: common to all transcripts, thus it is not considered for detecting variants; Region 2: specific to Transcript 1; Region 3: specific to Transcripts 2 and 3; Region 4: specific to Transcript 3; Region 5: specific to Transcript 1 and 2; Region 6: specific to Transcript 1.

EXAMPLE 5

Identification of cancer specific splice variants of genes over expressed in cancer

A search for EST supported (no mRNA) regions for genes of:

(i) known cancer markers

(ii) Genes shown to be over-expressed in cancer in published micro-array experiments.

Reliable EST supported-regions were defined as supported by minimum of one of the following:

(i) 3 spliced ESTs; or

(ii) 2 spliced ESTs from 2 libraries;

(iii) 10 unspliced ESTs from 2 libraries, or

(iv) 3 libraries.

Actual Marker Examples

The following examples relate to specific actual marker examples.

EXPERIMENTAL EXAMPLES SECTION

This Section relates to Examples describing experiments involving these sequences, and illustrative, non-limiting examples of methods, assays and uses thereof. The materials and experimental procedures are explained first, as all experiments used them as a basis for the work that was performed.

The markers of the present invention were tested with regard to their expression in various cancerous and non-cancerous tissue samples. A description of the samples used in the panel is provided in Table 2 below. A description of the samples used in the normal tissue panel is provided in Table 3 below. Tests were then performed as described in the "Materials and Experimental Procedures" sectionbelow.

Table 2: Tissue samples in testing panel

sample rename	Lot No.	source	pathology	Grade	gender/age
1-B-Adeno G1	A504117	Biochain	Adenocarcinoma	1	F/29
2-B-Adeno G1	A504118	Biochain	Adenocarcinoma	1	M/64
95-B-Adeno G1	A610063	Biochain	Adenocarcinoma	1	F/54
12-B-Adeno G2	A504119	Biochain	Adenocarcinoma	2	F/74
75-B-Adeno G2	A609217	Biochain	Adenocarcinoma	2	M/65
77-B-Adeno G2	A608301	Biochai	Adenocarcinom	2	M/44

		n	a		
13-B-Adeno G2-3	A504116	Biochain	Adenocarcinoma	2-3	M/64
89-B-Adeno G2-3	A609077	Biochain	Adenocarcinoma	2-3	M/62
76-B-Adeno G3	A609218	Biochain	Adenocarcinoma	3	M/57
94-B-Adeno G3	A610118	Biochain	Adenocarcinoma	3	M/68
3-CG-Adeno	CG-200	Ichilov	Adenocarcinoma		NA
14-CG- Adeno	CG-111	Ichilov	Adenocarcinoma		M/68
15-CG-Bronch adeno	CG-244	Ichilov	Bronchioloalveolar adenocarcinoma		M/74
45-B-Alvelous Adeno	A501221	Biochain	Alveolus carcinoma		F/50
44-B-Alvelous Adeno G2	A501123	Biochain	Alveolus carcinoma	2	F/61
19-B-Squamous G1	A408175	Biochain	Squamous carcinoma	1	M/78
16-B-Squamous G2	A409091	Biochain	Squamous carcinoma	2	F/68
17-B-Squamous G2	A503183	Biochain	Squamous carcinoma	2	M/57
21-B-Squamous G2	A503187	Biochain	Squamous carcinoma	2	M/52
78-B-Squamous G2	A607125	Biochain	Squamous Cell Carcinoma	2	M/62

80-B-Squamous G2	A609163	Biochain	Squamous Cell Carcinoma	2	M/74
18-B-Squamous G2-3	A503387	Biochain	Squamous Cell Carcinoma	2-3	M/63
81-B-Squamous G3	A609076	Biochain	Squamous Carcinoma	3	m/53
79-B-Squamous G3	A609018	Biochain	Squamous Cell Carcinoma	3	M/67
20-B-Squamous	A501121	Biochain	Squamous Carcinoma		M/64
22-B-Squamous	A503386	Biochain	Squamous Carcinoma		M/48
88-B-Squamous	A609219	Biochain	Squamous Cell Carcinoma		M/64
100-B-Squamous	A409017	Biochain	Squamous Carcinoma		M/64
23-CG-Squamous	CG-109 (1)	Ichilov	Squamous Carcinoma		M/65
24-CG-Squamous	CG-123	Ichilov	Squamous Carcinoma		M/76
25-CG-Squamous	CG-204	Ichilov	Squamous Carcinoma		M/72
87-B-Large cell G3	A609165	Biochain	Large Cell Carcinoma	3	F/47
38-B-Large cell	A504113	Biochain	Large cell		M/58
39-B-Large cell	A504114	Biochain	Large cell		F/35
82-B-Large cell	A609170	Biochain	Large Cell Neuroendocrine		M/68

			Carcinoma		
30-B-Small cell carci G3	A501389	Biochain	small cell	3	M/34
31-B-Small cell carci G3	A501390	Biochain	small cell	3	F/59
32-B-Small cell carci G3	A501391	Biochain	small cell	3	M/30
33-B-Small cell carci G3	A504115	Biochain	small cell	3	M
86-B-Small cell carci G3	A608032	Biochain	Small Cell Carcinoma	3	F/52
83-B-Small cell carci	A609162	Biochain	Small Cell Carcinoma		F/47
84-B-Small cell carci	A609167	Biochain	Small Cell Carcinoma		F/59
85-B-Small cell carci	A609169	Biochain	Small Cell Carcinoma		M/66
46-B-N M44	A501124	Biochain	Normal M44		F/61
47-B-N	A503205	Biochain	Normal PM		M/26
48-B-N	A503206	Biochain	Normal PM		M/44
49-B-N	A503384	Biochain	Normal PM		M/27
50-B-N	A503385	Biochain	Normal PM		M/28
90-B-N	A608152	Biochain	Normal (Pool 2) PM		pool 2
91-B-N	A607257	Biochai	Normal (Pool 2)		pool 2

		n	PM		
92-B-N	A503204	Biochain	Normal PM		m/28
93-Am-N	111P0103A	Ambion	Normal PM		F/61
96-Am-N	36853	Ambion	Normal PM		F/43
97-Am-N	36854	Ambion	Normal PM		M/46
98-Am-N	36855	Ambion	Normal PM		F/72
99-Am-N	36856	Ambion	Normal PM		M/31

Table 3: Tissue samples in normal panel:

	Lot no.	Source	Tissue	Pathology	Sex/Age
1-Am-Colon (C71)	071P10B	Ambion	Colon	PM	F/43
2-B-Colon (C69)	A411078	Biochain	Colon	PM-Pool of 10	M&F
3-Cl-Colon (C70)	1110101	Clontech	Colon	PM-Pool of 3	M&F
4-Am-Small Intestine	091P0201A	Ambion	Small Intestine	PM	M/75
5-B-Small Intestine	A501158	Biochain	Small Intestine	PM	M/63
6-B-Rectum	A605138	Biochain	Rectum	PM	M/25
7-B-Rectum	A610297	Biochain	Rectum	PM	M/24
8-B-Rectum	A610298	Biochain	Rectum	PM	M/27
9-Am-Stomach	110P04A	Ambion	Stomach	PM	M/16
10-B-Stomach	A501159	Biochain	Stomach	PM	M/24
11-B-Esophagus	A603814	Biochain	Esophagus	PM	M/26
12-B-Esophagus	A603813	Biochain	Esophagus	PM	M/41
13-Am-Pancreas	071P25C	Ambion	Pancreas	PM	M/25
14-CG-Pancreas	CG-255-2	Ichilov	Pancreas	PM	M/75

15-B-Lung	A409363	Biochain	Lung	PM	F/26
16-Am-Lung (L93)	111P0103A	Ambion	Lung	PM	F/61
17-B-Lung (L92)	A503204	Biochain	Lung	PM	M/28
18-Am-Ovary (O47)	061P43A	Ambion	Ovary	PM	F/16
19-B-Ovary (O48)	A504087	Biochain	Ovary	PM	F/51
20-B-Ovary (O46)	A504086	Biochain	Ovary	PM	F/41
21-Am-Cervix	101P0101A	Ambion	Cervix	PM	F/40
22-B-Cervix	A408211	Biochain	Cervix	PM	F/36
23-B-Cervix	A504089	Biochain	Cervix	PM-Pool of 5	M&F
24-B-Uterus	A411074	Biochain	Uterus	PM-Pool of 10	M&F
25-B-Uterus	A409248	Biochain	Uterus	PM	F/43
26-B-Uterus	A504090	Biochain	Uterus	PM-Pool of 5	M&F
27-B-Bladder	A501157	Biochain	Bladder	PM	M/29
28-Am-Bladder	071P02C	Ambion	Bladder	PM	M/20
29-B-Bladder	A504088	Biochain	Bladder	PM-Pool of 5	M&F
30-Am-Placenta	021P33A	Ambion	Placenta	PB	F/33
31-B-Placenta	A410165	Biochain	Placenta	PB	F/26
32-B-Placenta	A411073	Biochain	Placenta	PB-Pool of 5	M&F
33-B-Breast (B59)	A607155	Biochain	Breast	PM	F/36
34-Am-Breast (B63)	26486	Ambion	Breast	PM	F/43
35-Am-Breast (B64)	23036	Ambion	Breast	PM	F/57
36-Cl-Prostate (P53)	1070317	Clontech	Prostate	PB-Pool of 47	M&F
37-Am-Prostate (P42)	061P04A	Ambion	Prostate	PM	M/47
38-Am-Prostate (P59)	25955	Ambion	Prostate	PM	M/62
39-Am-Testis	111P0104A	Ambion	Testis	PM	M/25
40-B-Testis	A411147	Biochain	Testis	PM	M/74
41-Cl-Testis	1110320	Clontech	Testis	PB-Pool of 45	M&F
42-CG-Adrenal	CG-184-10	Ichilov	Adrenal	PM	F/81
43-B-Adrenal	A610374	Biochain	Adrenal	PM	F/83
44-B-Heart	A411077	Biochain	Heart	PB-Pool of 5	M&F

45-CG-Heart	CG-255-9	Ichilov	Heart	PM	M/75
46-CG-Heart	CG-227-1	Ichilov	Heart	PM	F/36
47-Am-Liver	081P0101A	Ambion	Liver	PM	M/64
48-CG-Liver	CG-93-3	Ichilov	Liver	PM	F/19
49-CG-Liver	CG-124-4	Ichilov	Liver	PM	F/34
50-Cl-BM	1110932	Clontech	Bone Marrow	PM-Pool of 8	M&F
51-CGEN-Blood	WBC#5	CGEN	Blood		M
52-CGEN-Blood	WBC#4	CGEN	Blood		M
53-CGEN-Blood	WBC#3	CGEN	Blood		M
54-CG-Spleen	CG-267	Ichilov	Spleen	PM	F/25
55-CG-Spleen	111P0106B	Ambion	Spleen	PM	M/25
56-CG-Spleen	A409246	Biochain	Spleen	PM	F/12
56-CG-Thymus	CG-98-7	Ichilov	Thymus	PM	F/28
58-Am-Thymus	101P0101A	Ambion	Thymus	PM	M/14
59-B-Thymus	A409278	Biochain	Thymus	PM	M/28
60-B-Thyroid	A610287	Biochain	Thyroid	PM	M/27
61-B-Thyroid	A610286	Biochain	Thyroid	PM	M/24
62-CG-Thyroid	CG-119-2	Ichilov	Thyroid	PM	F/66
63-Cl-Salivary Gland	1070319	Clontech	Salivary Gland	PM-Pool of 24	M&F
64-Am-Kidney	111P0101B	Ambion	Kidney	PM-Pool of 14	M&F
65-Cl-Kidney	1110970	Clontech	Kidney	PM-Pool of 14	M&F
66-B-Kidney	A411080	Biochain	Kidney	PM-Pool of 5	M&F
67-CG-Cerebellum	CG-183-5	Ichilov	Cerebellum	PM	M/74
68-CG-Cerebellum	CG-212-5	Ichilov	Cerebellum	PM	M/54
69-B-Brain	A411322	Biochain	Brain	PM	M/28
70-Cl-Brain	1120022	Clontech	Brain	PM-Pool of 2	M&F
71-B-Brain	A411079	Biochain	Brain	PM-Pool of 2	M&F
72-CG-Brain	CG-151-1	Ichilov	Brain	PM	F/86
73-Am-Skeletal Muscle	101P013A	Ambion	Skeletal Muscle	PM	F/28
74-Cl-Skeletal Muscle	1061038	Clontech	Skeletal Muscle	PM-Pool of 2	M&F

Materials and Experimental Procedures

RNA preparation – RNA was obtained from Clontech (Franklin Lakes, NJ USA 07417, www.clontech.com), BioChain Inst. Inc. (Hayward, CA 94545 USA www.biochain.com), ABS (Wilmington, DE 19801, USA, http://www.absbioreagents.com) or Ambion (Austin, TX 78744 USA, http://www.ambion.com). Alternatively, RNA was generated from tissue samples using TRI-Reagent (Molecular Research Center), according to Manufacturer's instructions. Tissue and RNA samples were obtained from patients or from postmortem. Total RNA samples were treated with DNaseI (Ambion) and purified using RNeasy columns (Qiagen).

RT PCR – Purified RNA (1 μg) was mixed with 150 ng Random Hexamer primers (Invitrogen) and 500 μM dNTP in a total volume of 15.6 μl. The mixture was incubated for 5 min at 65 °C and then quickly chilled on ice. Thereafter, 5 μl of 5X SuperscriptII first strand buffer (Invitrogen), 2.4μl 0.1M DTT and 40 units RNasin (Promega) were added, and the mixture was incubated for 10 min at 25 °C, followed by further incubation at 42 °C for 2 min. Then, 1 μl (200units) of SuperscriptII (Invitrogen) was added and the reaction (final volume of 25μl) was incubated for 50 min at 42 °C and then inactivated at 70 °C for 15min. The resulting cDNA was diluted 1:20 in TE buffer (10 mM Tris pH=8, 1 mM EDTA pH=8).

Real-Time RT-PCR analysis- cDNA (5μl), prepared as described above, was used as a template in Real-Time PCR reactions using the SYBR Green I assay (PE Applied Biosystem) with specific primers and UNG Enzyme (Eurogentech or ABI or Roche). The amplification was effected as follows: 50 °C for 2 min, 95 °C for 10 min, and then 40 cycles of 95 °C for 15sec, followed by 60 °C for 1 min. Detection was performed by using the PE Applied Biosystem SDS 7000. The cycle in which the reactions achieved a threshold level (Ct) of fluorescence was registered and was used to calculate the relative transcript quantity in the RT reactions. The relative quantity was calculated using the equation Q=efficiency^{-Ct}. The efficiency of the PCR reaction was calculated from a standard curve, created by using serial dilutions of several reverse transcription (RT) reactions. To minimize inherent differences in the RT reaction, the resulting relative quantities were normalized to the geometric mean of the relative quantities of several housekeeping (HSKP) genes. Schematic summary of quantitative real-time PCR

analysis is presented in Figure 3. As shown, the x-axis shows the cycle number. The C_T = Threshold Cycle point, which is the cycle that the amplification curve crosses the fluorescence threshold that was set in the experiment. This point is a calculated cycle number in which PCR product signal is above the background level (passive dye ROX) and still in the Geometric/Exponential phase (as shown, once the level of fluorescence crosses the measurement threshold, it has a geometrically increasing phase, during which measurements are most accurate, followed by a linear phase and a plateau phase; for quantitative measurements, the latter two phases do not provide accurate measurements). The y-axis shows the normalized reporter fluorescence. It should be noted that this type of analysis provides relative quantification.

The sequences of the housekeeping genes measured in all the examples in testing panel were as follows:

Ubiquitin (GenBank Accession No. BC000449)
Ubiquitin Forward primer (SEQ ID NO: 326): ATTTGGGTCGCGGTTCTTG
Ubiquitin Reverse primer (SEQ ID NO: 327): TGCCTTGACATTCTCGATGGT
Ubiquitin-amplicon (SEQ ID NO: 328)
ATTTGGGTCGCGGTTCTTGTTTGTGGATCGCTGTGATCGTCACTTGACAATGCAGAT
CTTCGTGAAGACTCTGACTGGTAAGACCATCACCCTCGAGG
TTGAGCCCAGTGACACCATCGAGAATGTCAAGGCA

SDHA (GenBank Accession No. NM_004168)
SDHA Forward primer (SEQ ID NO: 329): TGGGAACAAGAGGGCATCTG
SDHA Reverse primer (SEQ ID NO: 330): CCACCACTGCATCAAATTCATG
SDHA-amplicon (SEQ ID NO: 331):
TGGGAACAAGAGGGCATCTGCTAAAGTTTCAGATTCCATTTCTGCTCAGTATCCAGT
AGTGGATCATGAATTTGATGCAGTGGTGG

PBGD (GenBank Accession No. BC019323),
PBGD Forward primer (SEQ ID NO: 332): TGAGAGTGATTCGCGTGGG

PBGD Reverse primer (SEQ ID NO: 333): CCAGGGTACGAGGCTTTCAAT

PBGD-amplicon (SEQ ID NO: 334):

TGAGAGTGATTCGCGTGGGTACCCGCAAGAGCCAGCTTGCTCGCATACAGACGGAC AGTGTGGTGGCAACATTGAAAGCCTCGTACCCTGG

HPRT1 (GenBank Accession No. NM_000194),

HPRT1 Forward primer (SEQ ID NO: 1295): TGACACTGGCAAAACAATGCA

HPRT1 Reverse primer (SEQ ID NO: 1296): GGTCCTTTTC ACCAGCAAGCT

HPRT1-amplicon (SEQ ID NO: 1297):

TGACACTGGCAAAACAATGCAGACTTTGCTTTCCTTGGTCAGGCAGTATAATCCAA AGATGGTCAAGGTCGCAAGCTTGCTGGTGAAAAGGACC

The sequences of the housekeeping genes measured in all the examples on normal tissue samples panel were as follows:

RPL19 (GenBank Accession No. NM_000981),

RPL19 Forward primer (SEQ ID NO: 1298): TGGCAAGAAGAAGGTCTGGTTAG

RPL19 Reverse primer (SEQ ID NO: 1420): TGATCAGCCCATCTTTGATGAG

RPL19 –amplicon (SEQ ID NO: 1630):

TGGCAAGAAGAAGGTCTGGTTAGACCCCAATGAGACCAATGAAATCGCCAATGCCA ACTCCCGTCAGCAGATCCGGAAGCTCATCAAAGATGGGCTGATCA

TATA box (GenBank Accession No. NM_003194),

TATA box Forward primer (SEQ ID NO: 1631) : CGGTTTGCTGCGGTAATCAT

TATA box Reverse primer (SEQ ID NO: 1632): TTTCTTGCTGCCAGTCTGGAC

TATA box –amplicon (SEQ ID NO: 1633):

CGGTTTGCTGCGGTAATCATGAGGATAAGAGAGCCACGAACCACGGCACTGATTTT CAGTTCTGGGAAAATGGTGTGCACAGGAGCCAAGAGTGAAGAACAGTCCAGACTG GCAGCAAGAAA

Ubiquitin (GenBank Accession No. BC000449)

Ubiquitin Forward primer (SEQ ID NO: 326): ATTTGGGTCGCGGTTCTTG

Ubiquitin Reverse primer (SEQ ID NO: 327): TGCCTTGACATTCTCGATGGT

Ubiquitin-amplicon (SEQ ID NO: 328)
ATTTGGGTCGCGGTTCTTGTTTGTGGATCGCTGTGATCGTCACTTGACAATGCAGAT
CTTCGTGAAGACTCTGACTGGTAAGACCATCACCCTCGAGG
TTGAGCCCAGTGACACCATCGAGAATGTCAAGGCA
SDHA (GenBank Accession No. NM_004168)
SDHA Forward primer (SEQ ID NO: 329): TGGGAACAAGAGGGCATCTG
SDHA Reverse primer (SEQ ID NO: 330): CCACCACTGCATCAAATTCATG
SDHA-amplicon (SEQ ID NO: 331):
TGGGAACAAGAGGGCATCTGCTAAAGTTTCAGATTCCATTTCTGCTCAGTATCCAGT
AGTGGATCATGAATTTGATGCAGTGGTGG

Oligonucleotide-based micro-array experiment protocol-

Microarray fabrication

Microarrays (chips) were printed by pin deposition using the MicroGrid II MGII 600 robot from BioRobotics Limited (Cambridge, UK). 50-mer oligonucleotides target sequences were designed by Compugen Ltd (Tel-Aviv, IL) as described by A. Shoshan *et al*, "Optical technologies and informatics", Proceedings of SPIE. Vol 4266, pp. 86-95 (2001). The designed oligonucleotides were synthesized and purified by desalting with the Sigma-Genosys system (The Woodlands, TX, US) and all of the oligonucleotides were joined to a C6 amino-modified linker at the 5' end, or being attached directly to CodeLink slides (Cat #25-6700-01. Amersham Bioscience, Piscataway, NJ, US). The 50-mer oligonucleotides, forming the target sequences, were first suspended in Ultra-pure DDW (Cat # 01-866-1A Kibbutz Beit-Haemek, Israel) to a concentration of 50μM. Before printing the slides, the oligonucleotides were resuspended in 300mM sodium phosphate (pH 8.5) to final concentration of 150mM and printed at 35-40% relative humidity at 21°C.

Each slide contained a total of 9792 features in 32 subarrays. Of the se features, 4224 features were sequences of interest according to the present invention and negative controls that were printed in duplicate. An additional 288 features (96 target sequences printed in triplicate) contained housekeeping genes from Human Evaluation Library2, Compugen Ltd, Israel.

Another 384 features are E.coli spikes 1-6, which are oligos to E-Coli genes which are commercially available in the Array Control product (Array control- sense oligo spots, Ambion Inc. Austin, TX. Cat #1781, Lot #112K06).

Post-coupling processing of printed slides

After the spotting of the oligonucleotides to the glass (CodeLink) slides, the slides were incubated for 24 hours in a sealed saturated NaCl humidification chamber (relative humidity 70-75%).

Slides were treated for blocking of the residual reactive groups by incubating them in blocking solution at 50°C for 15 minutes (10ml/slide of buffer containing 0.1M Tris, 50mM ethanolamine, 0.1% SDS). The slides were then rinsed twice with Ultra-pure DDW (double distilled water). The slides were then washed with wash solution (10ml/slide. 4X SSC, 0.1% SDS)) at 50°C for 30 minutes on the shaker. The slides were then rinsed twice with Ultra-pure DDW, followed by drying by centrifugation for 3 minutes at 800 rpm.

Next, in order to assist in automatic operation of the hybridization protocol, the slides were treated with Ventana Discovery hybridization station barcode adhesives. The printed slides were loaded on a Bio-Optica (Milan, Italy) hematology staining device and were incubated for 10 minutes in 50ml of 3-Aminopropyl Triethoxysilane (Sigma A3648 lot #122K589). Excess fluid was dried and slides were then incubated for three hours in 20 mm/Hg in a dark vacuum desiccator (Pelco 2251, Ted Pella, Inc. Redding CA).

The following protocol was then followed with the Genisphere 900-RP (random primer), with mini elute columns on the Ventana Discovery HybStation™, to perform the microarray experiments. Briefly, the protocol was performed as described with regard to the instructions and information provided with the device itself. The protocol included cDNA synthesis and labeling. cDNA concentration was measured with the TBS-380 (Turner Biosystems. Sunnyvale, CA.) PicoFlour, which is used with the OliGreen ssDNA Quantitation reagent and kit.

Hybridization was performed with the Ventana Hybridization device, according to the provided protocols (Discovery Hybridization Station Tuscon AZ).

The slides were then scanned with GenePix 4000B dual laser scanner from Axon Instruments Inc, and analyzed by GenePix Pro 5.0 software.

Schematic summary of the oligonucleotide based microarray fabrication and the experimental flow is presented in Figures 4 and 5.

Briefly, as shown in Figure 4, DNA oligonucleotides at 25uM were deposited (printed) onto Amersham 'CodeLink' glass slides generating a well defined 'spot'. These slides are covered with a long-chain, hydrophilic polymer chemistry that creates an active 3-D surface that covalently binds the DNA oligonucleotides 5'-end via the C6-amine modification. This binding ensures that the full length of the DNA oligonucleotides is available for hybridization to the cDNA and also allows lower background, high sensitivity and reproducibility.

Figure 5 shows a schematic method for performing the microarray experiments. It should be noted that stages on the left-hand or right-hand side may optionally be performed in any order, including in parallel, until stage 4 (hybridization). Briefly, on the left-hand side, the target oligonucleotides are being spotted on a glass microscope slide (although optionally other materials could be used) to form a spotted slide (stage 1). On the right hand side, control sample RNA and cancer sample RNA are Cy3 and Cy5 labeled, respectively (stage 2), to form labeled probes. It should be noted that the control and cancer samples come from corresponding tissues (for example, normal prostate tissue and cancerous prostate tissue). Furthermore, the tissue from which the RNA was taken is indicated below in the specific examples of data for particular clusters, with regard to overexpression of an oligonucleotide from a "chip" (microarray), as for example "prostate" for chips in which prostate cancerous tissue and normal tissue were tested as described above. In stage 3, the probes are mixed. In stage 4, hybridization is performed to form a processed slide. In stage 5, the slide is washed and scanned to form an image file, followed by data analysis in stage 6.

The following clusters were found to be overexpressed in lung cancer:

W60282_PEA_1

F05068_PEA_1

H38804_PEA_1

HSENA78

T39971
(R00299)
H14624
Z41644_PEA_1
Z25299_PEA_2
HSSTROL3
HUMTREFAC_PEA_2
HSS100PCB
HSU33147_PEA_1
HUMCA1XIA
H61775
HUMGRP5E
HUMODCA
AA161187
R66178
D56406_PEA_1
M85491_PEA_1
Z21368_PEA_1
HUMCA1XIA
R20779
R38144_PEA_2
Z44808_PEA_1
HUMOSTRO_PEA_1_PEA_1
R11723_PEA_3
AI076020
T23580
M79217_PEA_1
M62096_PEA_1
M78076_PEA_1
T99080_PEA_4
T08446_PEA_1

R16276_PEA_1

The following clusters were found to be overexpressed in lung small cell cancer:

H61775

HUMGRP5E

M85491_PEA_1

Z44808_PEA_1

AA161187

R66178

HUMPHOSLIP_PEA_2

AI076020

T23580

M79217_PEA_1

M62096_PEA_1

M78076_PEA_1

T99080_PEA_4

T08446_PEA_1

The following clusters were found to be overexpressed in lung adenocarcinoma:

R00299

M85491_PEA_1

Z21368_PEA_1

HUMCA1XIA

AA161187

R66178

T11628_PEA_1

The following clusters were found to be overexpressed in lung squamous cell:

HUMODCA

R00299

D56406_PEA_1

Z44808_PEA_1

Z21368_PEA_1

HUMCA1XIA

AA161187

R66178

HUMCEA_PEA_1

R35137_PEA_1_PEA_1_PEA_1

DESCRIPTION FOR CLUSTER H61775

Cluster H61775 features 2 transcript(s) and 6 segment(s) of interest, the names for which are given in Tables 4 and 5, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 6.

Table 4 - Transcripts of interest

Transcript Name	Sequence ID No.
H61775_T21	1
H61775_T22	2

Table 5 - Segments of interest

Segment Name	Sequence ID No.
H61775_node_2	151
H61775_node_4	152
H61775_node_6	153
H61775_node_8	154
H61775_node_0	155
H61775_node_5	156

Table 6 - Proteins of interest

Protein Name	Sequence ID No.
H61775_P16	1281
H61775_P17	1282

Cluster H61775 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 6 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 6 and Table 7. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: brain malignant tumors and a mixture of malignant tumors from different tissues.

Table 7 - Normal tissue distribution

Name of Tissue	Number
bladder	0
brain	0
colon	0
epithelial	10
general	3
breast	8
muscle	0
ovary	0
pancreas	0

prostate	0
uterus	0

Table 8 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bladder	3.1e-01	3.8e-01	3.2e-01	2.5	4.6e-01	1.9
brain	8.8e-02	6.5e-02	1	3.5	4.1e-04	5.8
colon	5.6e-01	6.4e-01	1	1.1	1	1.1
epithelial	3.0e-02	1.3e-01	2.3e-02	2.1	3.2e-01	1.2
general	1.3e-06	4.9e-05	1.0e-07	6.3	1.5e-06	4.3
breast	4.7e-01	3.7e-01	3.3e-01	2.0	4.6e-01	1.6
muscle	2.3e-01	2.9e-01	1.5e-01	6.8	3.9e-01	2.6
ovary	3.8e-01	4.2e-01	1.5e-01	2.4	2.6e-01	1.9
pancreas	3.3e-01	4.4e-01	4.2e-01	2.4	5.3e-01	1.9
prostate	7.3e-01	7.8e-01	6.7e-01	1.5	7.5e-01	1.3
uterus	1.0e-01	2.6e-01	2.9e-01	2.6	5.1e-01	1.8

As noted above, contig H61775 features 2 transcript(s), which were listed in Table 3 above. A description of each variant protein according to the present invention is now provided.

Variant protein H61775_P16 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H61775_T21. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between H61775_P16 and Q9P2J2 (SEQ ID NO:1694):

1.An isolated chimeric polypeptide encoding for H61775_P16, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL

RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 11 - 93 of Q9P2J2, which also corresponds to amino acids 1 - 83 of H61775_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV corresponding to amino acids 84 - 152 of H61775_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of H61775_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV in H61775_P16.

Comparison report between H61775_P16 and AAQ88495 (SEQ ID NO:1695):

1.An isolated chimeric polypeptide encoding for H61775_P16, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 1 - 83 of AAQ88495, which also corresponds to amino acids 1 - 83 of H61775_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV corresponding to amino acids 84 - 152 of H61775_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of H61775_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCW RSSCSVTLQV in H61775_P16.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized

programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein H61775_P16 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 9, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H61775_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table9 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	I -> T	No
138	G -> R	No
34	G -> E	Yes
48	G -> R	No
91	R -> *	Yes

Variant protein H61775_P16 is encoded by the following transcript(s): H61775_T21, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H61775_T21 is shown in bold; this coding portion starts at position 261 and ends at position 716. The transcript also has the following SNPs as listed in Table 10 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H61775_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 10 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?

117	T -> C	Yes
200	T -> C	No
672	G -> C	No
222	T -> C	Yes
301	T -> C	No
361	G -> A	Yes
377	G -> A	No
400	-> C	No
402	G -> C	No
531	C -> T	Yes
566	T -> C	No

Variant protein H61775_P17 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H61775_T22. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between H61775_P17 and Q9P2J2:

1.An isolated chimeric polypeptide encoding for H61775_P17, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 11 - 93 of Q9P2J2, which also corresponds to amino acids 1 - 83 of H61775_P17.

Comparison report between H61775_P17 and AAQ88495:

1.An isolated chimeric polypeptide encoding for H61775_P17, comprising a first amino acid sequence being at least 90 % homologous to MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWL

RFGFLLPIFIQFGLYSPRIDPDYVG corresponding to amino acids 1 - 83 of AAQ88495, which also corresponds to amino acids 1 - 83 of H61775_P17.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein H61775_P17 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 11, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H61775_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 11 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	I -> T	No
34	G -> E	Yes
48	G -> R	No

Variant protein H61775_P17 is encoded by the following transcript(s): H61775_T22, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H61775_T22 is shown in bold; this coding portion starts at position 261 and ends at position 509. The transcript also has the following SNPs as listed in Table 12 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H61775_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 12 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
117	T -> C	Yes
200	T -> C	No
222	T -> C	Yes
301	T -> C	No
361	G -> A	Yes
377	G -> A	No
400	-> C	No
402	G -> C	No
596	T -> A	Yes

As noted above, cluster H61775 features 6 segment(s), which were listed in Table 4 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster H61775_node_2 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H61775_T21 and H61775_T22. Table 13 below describes the starting and ending position of this segment on each transcript.

Table 13 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T21	87	318
H61775_T22	87	318

Segment cluster H61775_node_4 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): H61775_T21 and H61775_T22. Table 14 below describes the starting and ending position of this segment on each transcript.

Table 14 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T21	319	507
H61775_T22	319	507

Segment cluster H61775_node_6 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H61775_T22. Table 15 below describes the starting and ending position of this segment on each transcript.

Table 15 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T22	515	715

Segment cluster H61775_node_8 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H61775_T21. Table 16 below describes the starting and ending position of this segment on each transcript.

Table 16 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T21	508	1205

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster H61775_node_0 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H61775_T21 and H61775_T22. Table 17 below describes the starting and ending position of this segment on each transcript.

Table 17 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T21	1	86
H61775_T22	1	86

Segment cluster H61775_node_5 according to the present invention can be found in the following transcript(s): H61775_T22. Table 18 below describes the starting and ending position of this segment on each transcript.

Table 18 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H61775_T22	508	514

Microarray (chip) data is also available for this gene as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (with regard to lung cancer), shown in Table 19.

Table 19 - Oligonucleotides related to this gene

Oligonucleotide name	Overexpressed in cancers	Chip reference
H61775_0_11_0	Lung cancer	Lung

Variant protein alignment to the previously known protein:

Sequence name: /tmp/Psw0RJLCti/aLAXQjXh07:Q9P2J2

Sequence documentation:

Alignment of: H61775_P16 x Q9P2J2 ..

Alignment segment 1/1:

Quality: 803.00

Escore: 0

Matching length: 83 Total length: 83

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       11 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 60
                .         .         .
       51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG                   83
          |||||||||||||||||||||||||||||||||
       61 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG                   93
```

Sequence name: /tmp/Psw0RJLCti/aLAXQjXh07:AAQ88495

Sequence documentation:

Alignment of: H61775_P16 x AAQ88495 ..

Alignment segment 1/1:

Quality: 803.00
Escore: 0
Matching length: 83 Total length: 83
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
                  .         .         .
       51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG                  83
          |||||||||||||||||||||||||||||||||
```

205

51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG 83

Sequence name: /tmp/naab8yR3GC/pSM412IL5o:Q9P2J2

Sequence documentation:

Alignment of: H61775_P17 x Q9P2J2 ..

Alignment segment 1/1:

Quality: 803.00

Escore: 0

Matching length: 83 Total length: 83

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       11 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 60
                .         .         .
```

206

```
51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG 83
   |||||||||||||||||||||||||||||||||
61 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG 93
```

Sequence name: /tmp/naab8yR3GC/pSM4l2IL5o:AAQ88495

Sequence documentation:

Alignment of: H61775_P17 x AAQ88495 ..

Alignment segment 1/1:

Quality: 803.00
Escore: 0
Matching length: 83 Total length: 83
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
          .         .         .         .         .
1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
  ||||||||||||||||||||||||||||||||||||||||||||||||||
```

207

```
  1 MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP 50
                    .          .          .
 51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG                  83
    ||||||||||||||||||||||||||||||||||
 51 PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVG                  83
```

Expression of immunoglobulin superfamily, member 9, H61775 transcripts which are detectable by amplicon as depicted in sequence name H61775seg8 in normal and cancerous lung tissues

Expression of immunoglobulin superfamily, member 9 transcripts detectable by or according to seg8 , H61775seg8 amplicon (SEQ ID NO: 1636) and H61775seg8F2 (SEQ ID NO: 1634) and H61775seg8R2 (SEQ ID NO: 1635) primers was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334, primers SEQ ID NOs 332 and 333), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297; primers SEQ ID NOs 1295 and 1296), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328, primers SEQ ID NOs 326 and 327) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331; primers SEQ ID NOs 329 and 330) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 7 is a histogram showing over expression of the above-indicated immunoglobulin superfamily, member 9 transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 5 fold over-expression, out of the total number of samples tested, is indicated in the bottom.

As is evident from Figure 7, the expression of immunoglobulin superfamily, member 9 transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than

in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99, Table 2 "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 11 out of 15 adenocarcinoma samples, 12 out of 16 squamous cell carcinoma samples, 1 out of 4 samples of large cell carcinoma samples and in 8 out of 8 small cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of immunoglobulin superfamily, member 9 transcripts detectable by the above amplicon in lung cancer samples versus the normal tissue samples was determined by T test as 6.5E-02. In adenocarcinoma, the minimum values were 7.62E-03 in squamous cell adenocarcinoma cancer and 1.5E-03 in small cell carcinoma.

Threshold of 5 fold overexpression was found to differentiate between cancer and normal samples with P value of 9.62E-04 in adenocarcinoma, 5.9E-04 in squamous cell carcinoma, and a threshold of 10 fold overexpression was found to differentiate between small cell adenocarcinoma cancer and normal samples with P value of 7.14E-05 as checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: H61775seg8F2 forward primer; and H61775seg8R2 reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: H61775seg8.

H61775seg8F2 (SEQ ID NO: 1634)
GAAGGCTCTTGTCACTTACTAGCCAT
H61775seg8R2 (SEQ ID NO: 1635)
TGTCACCATATTTAATCCTCCCAA
H61775seg8 (SEQ ID NO: 1636)
GAAGGCTCTTGTCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATTC
CTTCAGCTACAATGGAATTCTTGGGAGGATTAAATATGGTGACA

Expression of immunoglobulin superfamily, member 9, H61775 transcripts which are detectable by amplicon as depicted in sequence name H61775seg8 in different normal tissues.

Expression of immunoglobulin superfamily, member 9 transcripts detectable by or according to H61775 seg8 amplicon (SEQ ID NO: 1636) and H61775 seg8F2 (SEQ ID NO: 1634) and H61775 seg8R2 (SEQ ID NO: 1635) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20, Table 4, "Tissue sample in normal panel", above), to obtain a value of relative expression of each sample relative to median of the ovary samples.

H61775seg8F2 (SEQ ID NO: 1634)
GAAGGCTCTTGTCACTTACTAGCCAT
H61775seg8R2 (SEQ ID NO: 1635)
TGTCACCATATTTAATCCTCCCAA
H61775seg8 (SEQ ID NO: 1636)
GAAGGCTCTTGTCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATTC
CTTCAGCTACAATGGAATTCTTGGGAGGATTAAATATGGTGACA

The results are demonstrated in Figure 8, showing expression of immunoglobulin superfamily, member 9, H61775 transcripts, which are detectable by amplicon as depicted in sequence name H61775seg8, in different normal tissues.

DESCRIPTION FOR CLUSTER M85491

Cluster M85491 features 2 transcript(s) and 11 segment(s) of interest, the names for which are given in Tables 20 and 21, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 22.

Table 20 - Transcripts of interest

Transcript Name	Sequence ID No.
M85491_PEA_1_T16	3
M85491_PEA_1_T20	4

Table 21 - Segments of interest

Segment Name	Sequence ID No.
M85491_PEA_1_node_0	157
M85491_PEA_1_node_13	158
M85491_PEA_1_node_21	159
M85491_PEA_1_node_23	160
M85491_PEA_1_node_24	161
M85491_PEA_1_node_8	162
M85491_PEA_1_node_9	163
M85491_PEA_1_node_10	164
M85491_PEA_1_node_18	165
M85491_PEA_1_node_19	166
M85491_PEA_1_node_6	167

Table22 - Proteins of interest

Protein Name	Sequence ID No.
M85491_PEA_1_P13	1283
M85491_PEA_1_P14	1284

These sequences are variants of the known protein Ephrin type-B receptor 2 [precursor] (SwissProt accession identifier EPB2_HUMAN; known also according to the synonyms EC 2.7.1.112; Tyrosine-protein kinase receptor EPH-3; DRT; Receptor protein-tyrosine kinase HEK5; ERK), SEQ ID NO: 1417, referred to herein as the previously known protein.

Protein Ephrin type-B receptor 2 [precursor] is known or believed to have the following function(s): Receptor for members of the ephrin-B family. The sequence for protein Ephrin type-B receptor 2 [precursor] is given at the end of the application, as "Ephrin type-B receptor 2 [precursor] amino acid sequence" (SEQ ID NO:1417). Known polymorphisms for this sequence are as shown in Table 23.

Table 23 - Amino acid mutations for Known Protein

SNPposition(s) on amino acid sequence	Comment
671	A -> R. /FTId=VAR_004162.
1 - 20	MALRRLGAALLLLPLLAAVE -> MWVPVLALPVCTYA
923	E -> K
956	L -> V
958	V -> L
154	G -> D
476	K -> KQ
495 - 496	Missing
532	E -> D
568	R -> RR
589	M -> I
788	I -> F

853	S -> A

Protein Ephrin type-B receptor 2 [precursor] localization is believed to be Type I membrane protein.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: protein amino acid phosphorylation; transmembrane receptor protein tyrosine kinase signaling pathway; neurogenesis, which are annotation(s) related to Biological Process; protein tyrosine kinase; receptor; transmembrane-ephrin receptor; ATP binding; transferase, which are annotation(s) related to Molecular Function; and integral membrane protein, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster M85491 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 9 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 9 and Table 24. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Table24 - Normal tissue distribution

Name of Tissue	Number
Bladder	0
Bone	0

Brain	10
Colon	31
epithelial	10
General	12
Kidney	0
Liver	0
Lung	5
Breast	8
Muscle	5
Ovary	36
pancreas	10
Skin	0
Stomach	0

Table 25 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
Bladder	5.4e-01	6.0e-01	3.2e-01	2.5	4.6e-01	1.9
Bone	1	2.8e-01	1	1.0	7.0e-01	1.8
Brain	3.4e-01	3.6e-01	1.2e-01	2.9	1.8e-02	2.7
Colon	3.4e-02	5.7e-02	8.2e-02	2.8	2.0e-01	2.1
epithelial	1.7e-03	3.5e-03	2.0e-03	2.8	1.1e-02	2.2
General	4.8e-04	5.2e-04	6.7e-04	2.3	1.3e-03	1.9
Kidney	4.3e-01	3.7e-01	1	1.1	7.0e-01	1.5
Liver	1	4.5e-01	1	1.0	6.9e-01	1.5
Lung	2.2e-01	2.7e-01	6.9e-02	3.6	3.4e-02	3.6
Breast	8.2e-01	7.3e-01	6.9e-01	1.2	6.8e-01	1.2
Muscle	9.2e-01	4.8e-01	1	0.8	1.5e-01	3.2
Ovary	8.5e-01	7.3e-01	9.0e-01	0.7	6.7e-01	1.0
pancreas	5.5e-01	2.0e-01	6.7e-01	1.2	3.5e-01	1.8

Skin	2.9e-01	4.7e-01	1.4e-01	7.0	6.4e-01	1.6
Stomach	1.5e-01	3.2e-01	1	1.0	8.0e-01	1.3

As noted above, cluster M85491 features 2 transcript(s), which were listed in Table 20 above. These transcript(s) encode for protein(s) which are variant(s) of protein Ephrin type-B receptor 2 [precursor]. A description of each variant protein according to the present invention is now provided.

Variant protein M85491_PEA_1_P13 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M85491_PEA_1_T16. An alignment is given to the known protein (Ephrin type-B receptor 2 [precursor]) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M85491_PEA_1_P13 and EPB2_HUMAN:

1.An isolated chimeric polypeptide encoding for M85491_PEA_1_P13, comprising a first amino acid sequence being at least 90 % homologous to MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIR TYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYY EADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGGRVMKINTEVRSFGPVSRSGF YLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAEEVD VPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPIN SRTTSEGATNCVCRNGYYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDSG GREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQLGLTEPRIYISDLLAHTQYTFEIQAV NGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVDSITLSWSQPDQPNGVILDYEL QYYEK corresponding to amino acids 1 - 476 of EPB2_HUMAN, which also corresponds to amino acids 1 - 476 of M85491_PEA_1_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPIGWVLSPSPTSLRAPLPG corresponding to amino acids 477 - 496 of

M85491_PEA_1_P13, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M85491_PEA_1_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPIGWVLSPSPTSLRAPLPG in M85491_PEA_1_P13.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M85491_PEA_1_P13 is encoded by the following transcript(s): M85491_PEA_1_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M85491_PEA_1_T16 is shown in bold; this coding portion starts at position 143 and ends at position 1630. The transcript also has the following SNPs as listed in Table 26 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M85491_PEA_1_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 26 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
799	G -> A	Yes
1066	C -> T	Yes
1519	A -> G	Yes
1872	C -> T	Yes
2044	T -> C	Yes

2156	G -> A	Yes
2606	C -> A	Yes
2637	G -> C	Yes

Variant protein M85491_PEA_1_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M85491_PEA_1_T20. An alignment is given to the known protein (Ephrin type-B receptor 2 [precursor]) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M85491_PEA_1_P14 and EPB2_HUMAN:

1.An isolated chimeric polypeptide encoding for M85491_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIR TYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYY EADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGGRVMKINTEVRSFGPVSRSGF YLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAEEVD VPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCR corresponding to amino acids 1 - 270 of EPB2_HUMAN, which also corresponds to amino acids 1 - 270 of M85491_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ERQDLTMLSRLVLNSWPQMILPPQPPKVLEL corresponding to amino acids 271 - 301 of M85491_PEA_1_P14, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M85491_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ERQDLTMLSRLVLNSWPQMILPPQPPKVLEL in M85491_PEA_1_P14.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein M85491_PEA_1_P14 is encoded by the following transcript(s): M85491_PEA_1_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M85491_PEA_1_T20 is shown in bold; this coding portion starts at position 143 and ends at position 1045. The transcript also has the following SNPs as listed in Table 27 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M85491_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 27- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
799	G -> A	Yes
1135	T -> C	Yes
1160	T -> C	Yes
1172	A -> C	Yes
1176	T -> A	Yes

As noted above, cluster M85491 features 11 segment(s), which were listed in Table 21 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are

of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster M85491_PEA_1_node_0 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16 and M85491_PEA_1_T20. Table 28 below describes the starting and ending position of this segment on each transcript.

Table 28 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	1	203
M85491_PEA_1_T20	1	203

Segment cluster M85491_PEA_1_node_13 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T20. Table 29 below describes the starting and ending position of this segment on each transcript.

Table 29 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T20	954	1182

Segment cluster M85491_PEA_1_node_21 according to the present invention is supported by 18 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16. Table 30 below describes the starting and ending position of this segment on each transcript.

Table 30 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	1110	1445

Segment cluster M85491_PEA_1_node_23 according to the present invention is supported by 18 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16. Table 31 below describes the starting and ending position of this segment on each transcript.

Table 31 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	1446	1570

Segment cluster M85491_PEA_1_node_24 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16. Table 32 below describes the starting and ending position of this segment on each transcript.

Table 32- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	1571	2875

Segment cluster M85491_PEA_1_node_8 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16 and M85491_PEA_1_T20. Table 33 below describes the starting and ending position of this segment on each transcript.

Table 33 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	269	672
M85491_PEA_1_T20	269	672

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 34.

Table 34 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
M85491_0_14_0	lung malignant tumors	LUN

Segment cluster M85491_PEA_1_node_9 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16 and M85491_PEA_1_T20. Table 35 below describes the starting and ending position of this segment on each transcript.

Table 35 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	673	856
M85491_PEA_1_T20	673	856

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster M85491_PEA_1_node_10 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16 and M85491_PEA_1_T20. Table 36 below describes the starting and ending position of this segment on each transcript.

Table36 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	857	953

M85491_PEA_1_T20	857	953

Segment cluster M85491_PEA_1_node_18 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16. Table 37 below describes the starting and ending position of this segment on each transcript.

Table 37 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	954	1044

Segment cluster M85491_PEA_1_node_19 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16. Table 38 below describes the starting and ending position of this segment on each transcript.

Table 38 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	1045	1109

Segment cluster M85491_PEA_1_node_6 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M85491_PEA_1_T16 and M85491_PEA_1_T20. Table 39 below describes the starting and ending position of this segment on each transcript.

Table 39 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M85491_PEA_1_T16	204	268
M85491_PEA_1_T20	204	268

Variant protein alignment to the previously known protein:

Sequence name: /tmp/qfmsU9VtxS/DylcLC9j8v:EPB2_HUMAN

Sequence documentation:

Alignment of: M85491_PEA_1_P13 x EPB2_HUMAN ..

Alignment segment 1/1:

Quality: 4726.00

Escore: 0

Matching length: 476 Total length: 476

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYD 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYD 50
                .         .         .         .         .
       51 ENMNTIRTYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSI 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 ENMNTIRTYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSI 100
```

```
                .         .         .         .         .
101 PSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQV 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 PSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQV 150
                .         .         .         .         .
151 DLGGRVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRI 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 DLGGRVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRI 200
                .         .         .         .         .
201 IQNGAIFQETLSGAESTSLVAARGSCIANAEEVDVPIKLYCNGDGEWLVP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 IQNGAIFQETLSGAESTSLVAARGSCIANAEEVDVPIKLYCNGDGEWLVP 250
                .         .         .         .         .
251 IGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPINSRTTSEGA 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 IGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPINSRTTSEGA 300
                .         .         .         .         .
301 TNCVCRNGYYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDS 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 TNCVCRNGYYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDS 350
                .         .         .         .         .
351 GGREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQLGLTEPRIYISDLLA 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 GGREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQLGLTEPRIYISDLLA 400
                .         .         .         .         .
401 HTQYTFEIQAVNGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVD 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 HTQYTFEIQAVNGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVD 450
                .         .
451 SITLSWSQPDQPNGVILDYELQYYEK                          476
    ||||||||||||||||||||||||||
```

451 SITLSWSQPDQPNGVILDYELQYYEK 476

Sequence name: /tmp/rmnzuDbot6/GiHbjeU8iR:EPB2_HUMAN

Sequence documentation:

Alignment of: M85491_PEA_1_P14 x EPB2_HUMAN ..

Alignment segment 1/1:

Quality: 2673.00

Escore: 0

Matching length: 270 Total length: 270

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYD 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYD 50
                .         .         .         .         .
```

```
 51 ENMNTIRTYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSI 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 ENMNTIRTYQVCNVFESSQNNWLRTKFIRRRGAHRIHVEMKFSVRDCSSI 100
            .         .         .         .         .
101 PSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQV 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 PSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQV 150
            .         .         .         .         .
151 DLGGRVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRI 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 DLGGRVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRI 200
            .         .         .         .         .
201 IQNGAIFQETLSGAESTSLVAARGSCIANAEEVDVPIKLYCNGDGEWLVP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 IQNGAIFQETLSGAESTSLVAARGSCIANAEEVDVPIKLYCNGDGEWLVP 250
            .         .
251 IGRCMCKAGFEAVENGTVCR                               270
    ||||||||||||||||||||
251 IGRCMCKAGFEAVENGTVCR                               270
```

Expression of Ephrin type-B receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor EPH-3) M85491 transcripts which are detectable by amplicon as depicted in sequence name M85491seg24 in normal and cancerous lung tissues

Expression of Ephrin type-B receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor EPH-3) transcripts detectable by or according to seg24, M85491seg24 amplicon (SEQ ID NO: 1639) and M85491seg24F (SEQ ID NO: 1637) and M85491seg24R (SEQ ID NO: 1638) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon,

SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2 above, "Tissue samples in testing panel"), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 10 below is a histogram showing over expression of the above-indicated Ephrin type-B receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor EPH-3) transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained. The number and percentage of samples that exhibit at least 3 fold over-expression, out of the total number of samples tested, is indicated in the bottom.

As is evident from Figure 10, the expression of Ephrin type-B receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor EPH-3) transcripts detectable by the above ampliconin cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel".). Notably an over-expression of at least 3 fold was found in 9 out of 15 adenocarcinoma samples and in 4 out of 8 small cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

Threshold of 3 fold overexpression was found to differentiate between cancer and normal samples with P value of 7.42E-03 in adenocarcinoma and 5.69E-02 in small cell carcinoma as checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: *M85491seg24F* forward primer; and *M85491seg24R*reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: *M85491seg24*

M85491seg24F (SEQ ID NO: 1637) - GGCGTCTTTCTCCCTCTGAAC
M85491seg24R (SEQ ID NO: 1638) - GTCCCATTCTGGGTGCTGTG
M85491seg24 (SEQ ID NO: 1639)–
GGCGTCTTTCTCCCTCTGAACCTCAGTTTCCACCTGTGTCGAGTGTGGGTGAGACCC
CTCGCGGGGAGCTATGCAGGTTACGGAGAAAAGGCAGCACAGCACCCAGAATGGG
AC

Expression of Ephrin type-B receptor 2 precursor (EC 2.7.1.112) (Tyrosine-protein kinase receptor EPH-3)M85491 transcripts which are detectable by amplicon as depicted in sequence name M85491seg24 in different normal tissues

Expression of Ephrin type-B receptor 2 precursor transcripts detectable by or according to M85491 seg24 amplicon (SEQ ID NO: 1639) and M85491 seg24F (SEQ ID NO: 1637) and M85491 seg24R (SEQ ID NO: 1638) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the

median of the quantities of the lung samples (Sample Nos. 15-17, Table 2, "Tissue sample on normal panel", above), to obtain a value of relative expression of each sample relative to median of the lung samples.

M85491seg24F (SEQ ID NO: 1637) - GGCGTCTTTCTCCCTCTGAAC

M85491seg24R (SEQ ID NO: 1638) - GTCCCATTCTGGGTGCTGTG

M85491seg24 (SEQ ID NO: 1639)–

GGCGTCTTTCTCCCTCTGAACCTCAGTTTCCACCTGTGTCGAGTGTGGGTGAGACCC CTCGCGGGGAGCTATGCAGGTTACGGAGAAAAGGCAGCACAGCACCCAGAATGGG AC

The results are shown in Figure 11, demonstrating the expression of Ephrin type-B receptor 2 precursor (Tyrosine-protein kinase receptor EPH-3) M85491 transcripts which are detectable by amplicon as depicted in sequence name M85491seg24 in different normal tissues.

DESCRIPTION FOR CLUSTER T39971

Cluster T39971 features 4 transcript(s) and 28 segment(s) of interest, the names for which are given in Tables 40 and 41, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 42.

Table 40 - Transcripts of interest

Transcript Name	Sequence ID No.
T39971_T10	5
T39971_T12	6
T39971_T16	7
T39971_T5	8

Table 41 - Segments of interest

Segment Name	Sequence ID No.
T39971_node_0	168

T39971_node_18	169
T39971_node_21	170
T39971_node_22	171
T39971_node_23	172
T39971_node_31	173
T39971_node_33	174
T39971_node_7	175
T39971_node_1	176
T39971_node_10	177
T39971_node_11	178
T39971_node_12	179
T39971_node_15	180
T39971_node_16	181
T39971_node_17	182
T39971_node_26	183
T39971_node_27	184
T39971_node_28	185
T39971_node_29	186
T39971_node_3	187
T39971_node_30	188
T39971_node_34	189
T39971_node_35	190
T39971_node_36	191
T39971_node_4	192
T39971_node_5	193
T39971_node_8	194
T39971_node_9	195

Table 42 - Proteins of interest

Protein Name	Sequence ID No.
T39971_P6	1285
T39971_P9	1286
T39971_P11	1287
T39971_P12	1288

These sequences are variants of the known protein Vitronectin precursor (SwissProt accession identifier VTNC_HUMAN; known also according to the synonyms Serum spreading factor; S-protein; V75), SEQ ID NO: 1418, referred to herein as the previously known protein.

Protein Vitronectin precursor is known or believed to have the following function(s): Vitronectin is a cell adhesion and spreading factor found in serum and tissues. Vitronectin interacts with glycosaminoglycans and proteoglycans. Is recognized by certain members of the integrin family and serves as a cell-to-substrate adhesion molecule. Inhibitor of the membrane-damaging effect of the terminal cytolytic complement pathway. The sequence for protein Vitronectin precursor is given at the end of the application, as "Vitronectin precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 4.

Table 43 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
122	A -> S. /FTId=VAR_012983.
268	R -> Q. /FTId=VAR_012984.
400	T -> M. /FTId=VAR_012985.
50	C -> N
225	S -> N
366	A -> T

Protein Vitronectin precursor localization is believed to be Extracellular.

The previously known protein also has the following indication(s) and/or potential therapeutic use(s): Cancer, melanoma. It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct

therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Alphavbeta3 integrin antagonist; Apoptosis agonist. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Anticancer.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: immune response; cell adhesion, which are annotation(s) related to Biological Process; protein binding; heparin binding, which are annotation(s) related to Molecular Function; and extracellular space, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster T39971 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 12 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 12 and Table 44. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: liver cancer, lung malignant tumors and pancreas carcinoma.

Table 44 - Normal tissue distribution

Name of Tissue	Number
adrenal	60
bladder	0

Bone	0
Brain	9
Colon	0
epithelial	79
general	29
Liver	2164
Lung	0
Lymph nodes	0
Breast	0
pancreas	0
prostate	0
Skin	0
Uterus	0

Table 45 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	6.9e-01	7.4e-01	2.0e-02	2.3	5.3e-02	1.8
bladder	5.4e-01	6.0e-01	5.6e-01	1.8	6.8e-01	1.5
Bone	1	6.7e-01	1	1.0	7.0e-01	1.4
Brain	8.0e-01	8.6e-01	3.0e-01	1.9	5.3e-01	1.2
Colon	4.2e-01	4.8e-01	7.0e-01	1.6	7.7e-01	1.4
epithelial	6.6e-01	5.7e-01	1.0e-01	0.8	8.7e-01	0.6
general	5.1e-01	3.8e-01	9.2e-08	1.6	8.3e-04	1.3
Liver	1	6.7e-01	2.3e-03	0.3	1	0.2
Lung	2.4e-01	9.1e-02	1.7e-01	4.3	8.1e-03	5.0
Lymph nodes	1	5.7e-01	1	1.0	5.8e-01	2.3
Breast	1	6.7e-01	1	1.0	8.2e-01	1.2
pancreas	9.5e-02	1.8e-01	1.5e-11	6.5	8.2e-09	4.6
prostate	7.3e-01	6.0e-01	6.7e-01	1.5	5.6e-01	1.7

Skin	1	4.4e-01	1	1.0	6.4e-01	1.6
Uterus	5.0e-01	2.6e-01	1	1.1	8.0e-01	1.4

As noted above, cluster T39971 features 4 transcript(s), which were listed in Table 40 above. These transcript(s) encode for protein(s) which are variant(s) of protein Vitronectin precursor. A description of each variant protein according to the present invention is now provided.

Variant protein T39971_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T39971_T5. An alignment is given to the known protein (Vitronectin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T39971_P6 and VTNC_HUMAN:

1.An isolated chimeric polypeptide encoding for T39971_P6, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKG corresponding to amino acids 1 - 276 of VTNC_HUMAN, which also corresponds to amino acids 1 - 276 of T39971_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TQGVVGD corresponding to amino acids 277 - 283 of T39971_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of T39971_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TQGVVGD in T39971_P6.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein T39971_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 46, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table46 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
122	A -> S	Yes
145	G ->	No
268	R -> Q	Yes
280	V -> A	Yes
180	C ->	No
180	C -> W	No
192	Y ->	No
209	A ->	No
211	T ->	No
267	G ->	No
267	G -> A	No
268	R ->	No

Variant protein T39971_P6 is encoded by the following transcript(s): T39971_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T39971_T5 is shown in bold; this coding portion starts at position 756 and ends at position

1604. The transcript also has the following SNPs as listed in Table47 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 47 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
417	G -> C	Yes
459	T -> C	Yes
1387	C ->	No
1406	-> A	No
1406	-> G	No
1555	G ->	No
1555	G -> C	No
1558	G ->	No
1558	G -> A	Yes
1594	T -> C	Yes
1642	T -> C	Yes
1770	C -> T	Yes
529	G -> T	Yes
1982	A -> G	No
2007	G ->	No
2029	T -> C	No
2094	T -> C	No
2117	C -> G	No
2123	C -> T	Yes
2152	C -> T	Yes
2182	G -> T	No

2185	A -> C	No
2297	T -> C	Yes
1119	G -> T	Yes
2411	G ->	No
2411	G -> T	No
2487	T -> C	Yes
1188	G ->	No
1295	C ->	No
1295	C -> G	No
1324	-> T	No
1331	C ->	No
1381	C ->	No

Variant protein T39971_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T39971_T10. An alignment is given to the known protein (Vitronectin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T39971_P9 and VTNC_HUMAN:

1.An isolated chimeric polypeptide encoding for T39971_P9, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRT corresponding to amino acids 1 - 325 of

VTNC_HUMAN, which also corresponds to amino acids 1 - 325 of T39971_P9, and a second amino acid sequence being at least 90 % homologous to SGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRATWLSLFSSEESNLGA NNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLRTRRVDTVDPPYPRSIAQYWLGC PAPGHL corresponding to amino acids 357 - 478 of VTNC_HUMAN, which also corresponds to amino acids 326 - 447 of T39971_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of T39971_P9, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TS, having a structure as follows: a sequence starting from any of amino acid numbers 325-x to 325; and ending at any of amino acid numbers 326 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T39971_P9 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 48, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table48 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
122	A -> S	Yes

145	G ->	No
268	R -> Q	Yes
328	M -> T	No
350	S -> P	No
369	T -> M	Yes
379	S -> I	No
380	N -> T	No
180	C ->	No
180	C -> W	No
192	Y ->	No
209	A ->	No
211	T ->	No
267	G ->	No
267	G -> A	No
268	R ->	No

Variant protein T39971_P9 is encoded by the following transcript(s): T39971_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T39971_T10 is shown in bold; this coding portion starts at position 756 and ends at position 2096. The transcript also has the following SNPs as listed in Table 49 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 49 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
417	G -> C	Yes
459	T -> C	Yes
1387	C ->	No

1406	-> A	No
1406	-> G	No
1555	G ->	No
1555	G -> C	No
1558	G ->	No
1558	G -> A	Yes
1738	T -> C	No
1803	T -> C	No
1826	C -> G	No
529	G -> T	Yes
1832	C -> T	Yes
1861	C -> T	Yes
1891	G -> T	No
1894	A -> C	No
2006	T -> C	Yes
2120	G ->	No
2120	G -> T	No
2196	T -> C	Yes
1119	G -> T	Yes
1188	G ->	No
1295	C ->	No
1295	C -> G	No
1324	-> T	No
1331	C ->	No
1381	C ->	No

Variant protein T39971_P11 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T39971_T12. An alignment is given to the known protein (Vitronectin precursor) at the end of the application.

One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T39971_P11 and VTNC_HUMAN:

1.An isolated chimeric polypeptide encoding for T39971_P11, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRTS corresponding to amino acids 1 - 326 of VTNC_HUMAN, which also corresponds to amino acids 1 - 326 of T39971_P11, and a second amino acid sequence being at least 90 % homologous to DKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL corresponding to amino acids 442 - 478 of VTNC_HUMAN, which also corresponds to amino acids 327 - 363 of T39971_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of T39971_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise SD, having a structure as follows: a sequence starting from any of amino acid numbers 326-x to 326; and ending at any of amino acid numbers 327 + ((n-2) - x), in which x varies from 0 to n-2.

Comparison report between T39971_P11 and Q9BSH7 (SEQ ID NO:1696):

1.An isolated chimeric polypeptide encoding for T39971_P11, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC

KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGV LDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEE CEGSSLSAVFEHFAMMQRDSWEDIFELLFWGRTS corresponding to amino acids 1 - 326 of Q9BSH7, which also corresponds to amino acids 1 - 326 of T39971_P11, and a second amino acid sequence being at least 90 % homologous to DKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL corresponding to amino acids 442 - 478 of Q9BSH7, which also corresponds to amino acids 327 - 363 of T39971_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of T39971_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise SD, having a structure as follows: a sequence starting from any of amino acid numbers 326-x to 326; and ending at any of amino acid numbers 327 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein T39971_P11 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 50, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table50 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
122	A -> S	Yes
145	G ->	No
268	R -> Q	Yes
180	C ->	No
180	C -> W	No
192	Y ->	No
209	A ->	No
211	T ->	No
267	G ->	No
267	G -> A	No
268	R ->	No

Variant protein T39971_P11 is encoded by the following transcript(s): T39971_T12, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T39971_T12 is shown in bold; this coding portion starts at position 756 and ends at position 1844. The transcript also has the following SNPs as listed in Table 51 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 51 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
417	G -> C	Yes
459	T -> C	Yes
1387	C ->	No
1406	-> A	No

1406	-> G	No
1555	G ->	No
1555	G -> C	No
1558	G ->	No
1558	G -> A	Yes
1754	T -> C	Yes
1868	G ->	No
1868	G -> T	No
529	G -> T	Yes
1944	T -> C	Yes
1119	G -> T	Yes
1188	G ->	No
1295	C ->	No
1295	C -> G	No
1324	-> T	No
1331	C ->	No
1381	C ->	No

Variant protein T39971_P12 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T39971_T16. An alignment is given to the known protein (Vitronectin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T39971_P12 and VTNC_HUMAN:

1.An isolated chimeric polypeptide encoding for T39971_P12, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC

KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFK corresponding to amino acids 1 - 223 of VTNC_HUMAN, which also corresponds to amino acids 1 - 223 of T39971_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPGAVGQGRKHLGRV corresponding to amino acids 224 - 238 of T39971_P12, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of T39971_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPGAVGQGRKHLGRV in T39971_P12.

Comparison report between T39971_P12 and Q9BSH7:

1.An isolated chimeric polypeptide encoding for T39971_P12, comprising a first amino acid sequence being at least 90 % homologous to MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAEC KPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPV LKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEELCSGKPFDAFTDLKNGSLFAFR GQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFK corresponding to amino acids 1 - 223 of Q9BSH7, which also corresponds to amino acids 1 - 223 of T39971_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VPGAVGQGRKHLGRV corresponding to amino acids 224 - 238 of T39971_P12, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of T39971_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VPGAVGQGRKHLGRV in T39971_P12.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T39971_P12 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 52, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T39971_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 52 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
122	A -> S	Yes
145	G ->	No
180	C ->	No
180	C -> W	No
192	Y ->	No
209	A ->	No
211	T ->	No

Variant protein T39971_P12 is encoded by the following transcript(s): T39971_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T39971_T16 is shown in bold; this coding portion starts at position 756 and ends at position 1469. The transcript also has the following SNPs as listed in Table 53 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein

T39971_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 53 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
417	G -> C	Yes
459	T -> C	Yes
1387	C ->	No
1406	-> A	No
1406	-> G	No
529	G -> T	Yes
1119	G -> T	Yes
1188	G ->	No
1295	C ->	No
1295	C -> G	No
1324	-> T	No
1331	C ->	No
1381	C ->	No

As noted above, cluster T39971 features 28 segment(s), which were listed in Table 41 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster T39971_node_0 according to the present invention is supported by 76 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 54 below describes the starting and ending position of this segment on each transcript.

Table 54 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1	810
T39971_T12	1	810
T39971_T16	1	810
T39971_T5	1	810

Segment cluster T39971_node_18 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T16. Table 55 below describes the starting and ending position of this segment on each transcript.

Table 55 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T16	1425	1592

Segment cluster T39971_node_21 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 56 below describes the starting and ending position of this segment on each transcript.

Table56 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1425	1581
T39971_T12	1425	1581
T39971_T5	1425	1581

Segment cluster T39971_node_22 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): T39971_T5. Table 57 below describes the starting and ending position of this segment on each transcript.

Table57 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T5	1582	1779

Segment cluster T39971_node_23 according to the present invention is supported by 101 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 58 below describes the starting and ending position of this segment on each transcript.

Table 58 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1582	1734
T39971_T12	1582	1734
T39971_T5	1780	1932

Segment cluster T39971_node_31 according to the present invention is supported by 94 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10 and T39971_T5. Table 59 below describes the starting and ending position of this segment on each transcript.

Table 59 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1847	1986
T39971_T5	2138	2277

Segment cluster T39971_node_33 according to the present invention is supported by 77 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 60 below describes the starting and ending position of this segment on each transcript.

Table 60 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1987	2113
T39971_T12	1735	1861
T39971_T5	2278	2404

Segment cluster T39971_node_7 according to the present invention is supported by 87 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 61 below describes the starting and ending position of this segment on each transcript.

Table 61 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	940	1162
T39971_T12	940	1162
T39971_T16	940	1162
T39971_T5	940	1162

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster T39971_node_1 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 62 below describes the starting and ending position of this segment on each transcript.

Table 62 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	811	819
T39971_T12	811	819
T39971_T16	811	819
T39971_T5	811	819

Segment cluster T39971_node_10 according to the present invention is supported by 77 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 63 below describes the starting and ending position of this segment on each transcript.

Table 63 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1189	1232
T39971_T12	1189	1232
T39971_T16	1189	1232
T39971_T5	1189	1232

Segment cluster T39971_node_11 according to the present invention is supported by 79 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 64 below describes the starting and ending position of this segment on each transcript.

Table 64 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1233	1270
T39971_T12	1233	1270
T39971_T16	1233	1270
T39971_T5	1233	1270

Segment cluster T39971_node_12 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 65 below describes the starting and ending position of this segment on each transcript.

Table 65 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1271	1284
T39971_T12	1271	1284
T39971_T16	1271	1284
T39971_T5	1271	1284

Segment cluster T39971_node_15 according to the present invention is supported by 79 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 66 below describes the starting and ending position of this segment on each transcript.

Table 66 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1285	1316
T39971_T12	1285	1316
T39971_T16	1285	1316
T39971_T5	1285	1316

Segment cluster T39971_node_16 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 67 below describes the starting and ending position of this segment on each transcript.

Table 67 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1317	1340
T39971_T12	1317	1340
T39971_T16	1317	1340
T39971_T5	1317	1340

Segment cluster T39971_node_17 according to the present invention is supported by 86 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 68 below describes the starting and ending position of this segment on each transcript.

Table 68 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1341	1424
T39971_T12	1341	1424
T39971_T16	1341	1424
T39971_T5	1341	1424

Segment cluster T39971_node_26 according to the present invention is supported by 85 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T5. Table 69 below describes the starting and ending position of this segment on each transcript.

Table 69 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T5	1933	1974

Segment cluster T39971_node_27 according to the present invention is supported by 90 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T5. Table 70 below describes the starting and ending position of this segment on each transcript.

Table 70 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T5	1975	2025

Segment cluster T39971_node_28 according to the present invention can be found in the following transcript(s): T39971_T10 and T39971_T5. Table 71 below describes the starting and ending position of this segment on each transcript.

Table 71 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1735	1743
T39971_T5	2026	2034

Segment cluster T39971_node_29 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10 and T39971_T5. Table 72 below describes the starting and ending position of this segment on each transcript.

Table 72 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1744	1838
T39971_T5	2035	2129

Segment cluster T39971_node_3 according to the present invention is supported by 78 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 73 below describes the starting and ending position of this segment on each transcript.

Table 73 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	820	861
T39971_T12	820	861
T39971_T16	820	861
T39971_T5	820	861

Segment cluster T39971_node_30 according to the present invention can be found in the following transcript(s): T39971_T10 and T39971_T5. Table 74 below describes the starting and ending position of this segment on each transcript.

Table 74 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1839	1846
T39971_T5	2130	2137

Segment cluster T39971_node_34 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 75 below describes the starting and ending position of this segment on each transcript.

Table 75 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	2114	2120
T39971_T12	1862	1868
T39971_T5	2405	2411

Segment cluster T39971_node_35 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 76 below describes the starting and ending position of this segment on each transcript.

Table 76 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	2121	2137
T39971_T12	1869	1885
T39971_T5	2412	2428

Segment cluster T39971_node_36 according to the present invention is supported by 51 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12 and T39971_T5. Table 77 below describes the starting and ending position of this segment on each transcript.

Table 77 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	2138	2199
T39971_T12	1886	1947
T39971_T5	2429	2490

Segment cluster T39971_node_4 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 78 below describes the starting and ending position of this segment on each transcript.

Table 78 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	862	881

T39971_T12	862	881
T39971_T16	862	881
T39971_T5	862	881

Segment cluster T39971_node_5 according to the present invention is supported by 80 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 79 below describes the starting and ending position of this segment on each transcript.

Table 79 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	882	939
T39971_T12	882	939
T39971_T16	882	939
T39971_T5	882	939

Segment cluster T39971_node_8 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 80 below describes the starting and ending position of this segment on each transcript.

Table 80 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1163	1168
T39971_T12	1163	1168
T39971_T16	1163	1168
T39971_T5	1163	1168

Segment cluster T39971_node_9 according to the present invention can be found in the following transcript(s): T39971_T10, T39971_T12, T39971_T16 and T39971_T5. Table 81 below describes the starting and ending position of this segment on each transcript.

Table 81 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T39971_T10	1169	1188
T39971_T12	1169	1188
T39971_T16	1169	1188
T39971_T5	1169	1188

Variant protein alignment to the previously known protein:

Sequence name: /tmp/xkraCL2OcZ/43L7YcPH7x:VTNC_HUMAN

Sequence documentation:

Alignment of: T39971_P6 x VTNC_HUMAN ..

Alignment segment 1/1:

Quality: 2774.00

Escore: 0

Matching length: 278 Total length: 278

Matching Percent Similarity: 99.64 Matching Percent Identity: 99.64

Total Percent Similarity: 99.64 Total Percent Identity: 99.64

Gaps: 0

Alignment:

```
                 .         .         .         .         .
       1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                 .         .         .         .         .
      51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
                 .         .         .         .         .
     101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
                 .         .         .         .         .
     151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
                 .         .         .         .         .
     201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
                 .         .
     251 PDNVDAALALPAHSYSGRERVYFFKGTQ                       278
         |||||||||||||||||||||||||| |
     251 PDNVDAALALPAHSYSGRERVYFFKGKQ                       278
```

Sequence name: /tmp/X4DeeuSlB4/yMubSR5FPs:VTNC_HUMAN

Sequence documentation:

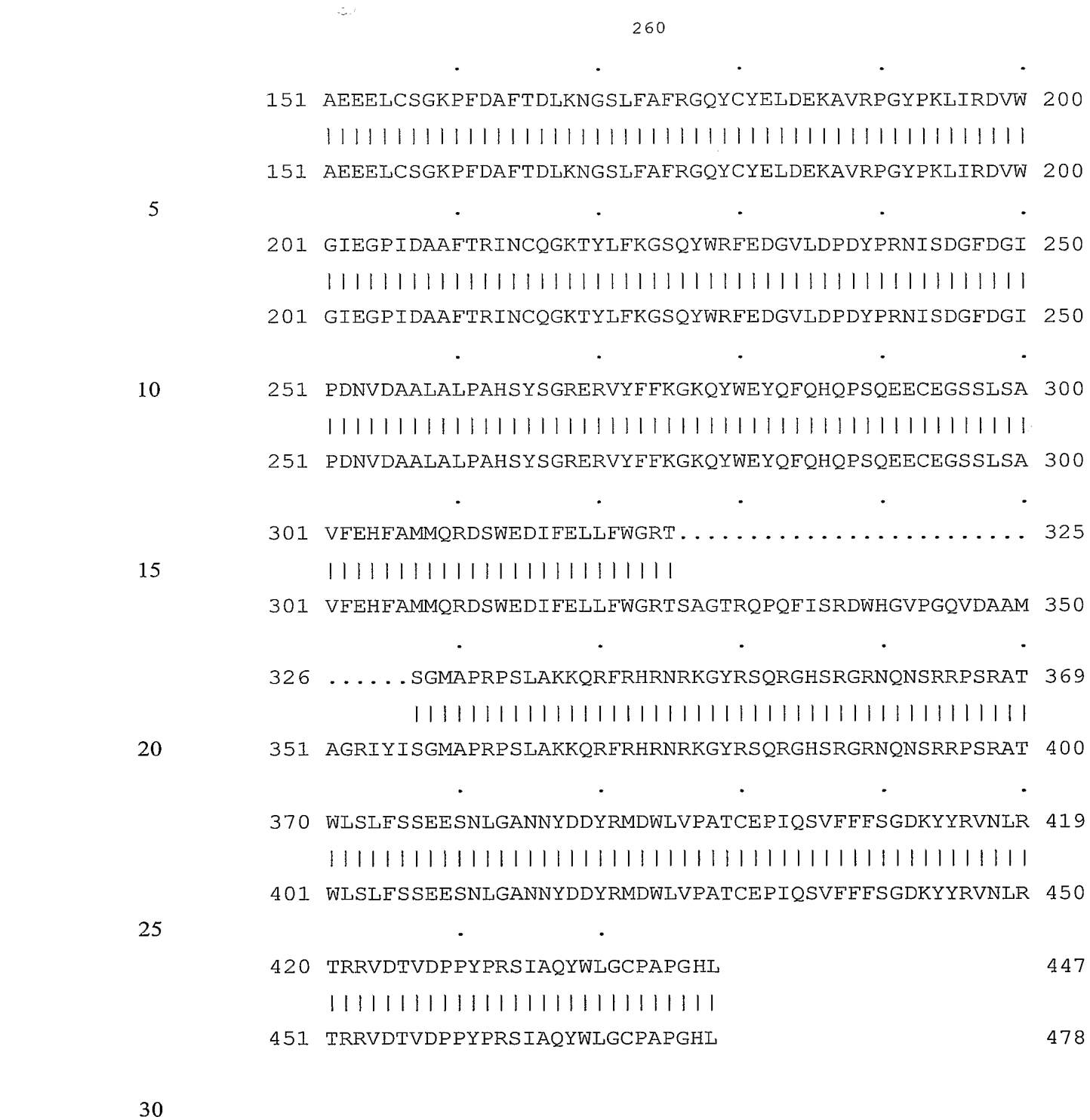
Alignment of: T39971_P9 x VTNC_HUMAN ..

Alignment segment 1/1:

Quality: 4430.00
Escore: 0
Matching length: 447 Total length: 478
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 93.51 Total Percent Identity: 93.51
Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                .         .         .         .         .
     51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
                .         .         .         .         .
    101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
```


260
151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
301 VFEHFAMMQRDSWEDIFELLFWGRT..... 325
301 VFEHFAMMQRDSWEDIFELLFWGRTSAGTRQPQFISRDWHGVPGQVDAAM 350
326 ......SGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRAT 369
351 AGRIYISGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRAT 400
370 WLSLFSSEESNLGANNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLR 419
401 WLSLFSSEESNLGANNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLR 450
420 TRRVDTVDPPYPRSIAQYWLGCPAPGHL 447
451 TRRVDTVDPPYPRSIAQYWLGCPAPGHL 478

Sequence name: /tmp/jvp1VtnxNy/wxNSeFVZZw:VTNC_HUMAN

Sequence documentation:

Alignment of: T39971_P11 x VTNC_HUMAN ..

Alignment segment 1/1:

Quality: 3576.00

Escore: 0

Matching length: 363 Total length: 478

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 75.94 Total Percent Identity: 75.94

Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                .         .         .         .         .
       51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
                .         .         .         .         .
```

```
101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
          .         .         .         .         .
151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
          .         .         .         .         .
201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
          .         .         .         .         .
251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
          .         .         .         .         .
301 VFEHFAMMQRDSWEDIFELLFWGRTS..................... 326
    |||||||||||||||||||||||||
301 VFEHFAMMQRDSWEDIFELLFWGRTSAGTRQPQFISRDWHGVPGQVDAAM 350
          .         .         .         .         .
326 ..................................................  326

351 AGRIYISGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRAT 400
          .         .         .         .         .
327 .....................................DKYYRVNLR 335
                                         |||||||||
401 WLSLFSSEESNLGANNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLR 450
          .         .
336 TRRVDTVDPPYPRSIAQYWLGCPAPGHL                      363
    ||||||||||||||||||||||||||||
451 TRRVDTVDPPYPRSIAQYWLGCPAPGHL                      478
```

Sequence name: /tmp/jvp1VtnxNy/wxNSeFVZZw:Q9BSH7

Sequence documentation:

Alignment of: T39971_P11 x Q9BSH7 ..

Alignment segment 1/1:

Quality: 3576.00
Escore: 0
Matching length: 363 Total length: 478
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 75.94 Total Percent Identity: 75.94
Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                .         .         .         .         .
       51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
               .         .         .         .         .
  101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
               .         .         .         .         .
  151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
               .         .         .         .         .
  201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 GIEGPIDAAFTRINCQGKTYLFKGSQYWRFEDGVLDPDYPRNISDGFDGI 250
               .         .         .         .         .
  251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  251 PDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSA 300
               .         .         .         .         .
  301 VFEHFAMMQRDSWEDIFELLFWGRTS........................ 326
      ||||||||||||||||||||||||||
  301 VFEHFAMMQRDSWEDIFELLFWGRTSAGTRQPQFISRDWHGVPGQVDAAM 350
               .         .         .         .         .
  326 .................................................. 326

  351 AGRIYISGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRAM 400
               .         .         .         .         .
  327 .........................................DKYYRVNLR 335
                                               |||||||||
  401 WLSLFSSEESNLGANNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLR 450
               .         .
```

```
                                    265
        336 TRRVDTVDPPYPRSIAQYWLGCPAPGHL                    363
            |||||||||||||||||||||||||||||
        451 TRRVDTVDPPYPRSIAQYWLGCPAPGHL                    478
```

Sequence name: /tmp/fgebv7ir4i/48bTBMziJ0:VTNC_HUMAN

Sequence documentation:

Alignment of: T39971_P12 x VTNC_HUMAN ..

Alignment segment 1/1:

Quality: 2237.00
Escore: 0
Matching length: 223 Total length: 223
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
          1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
          1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                      .         .         .         .         .
         51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
         51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
                      .         .         .         .         .
        101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
        101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
                      .         .         .         .         .
        151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
        151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
                      .         .
        201 GIEGPIDAAFTRINCQGKTYLFK                             223
            |||||||||||||||||||||||
        201 GIEGPIDAAFTRINCQGKTYLFK                             223
```

Sequence name: /tmp/fgebv7ir4i/48bTBMziJ0:Q9BSH7

Sequence documentation:

Alignment of: T39971_P12 x Q9BSH7 ..

Alignment segment 1/1:

Quality: 2237.00

Escore: 0

Matching length: 223 Total length: 223

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSC 50
                .         .         .         .         .
     51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CTDYTAECKPQVTRGDVFTMPEDEYTVYDDGEEKNNATVHEQVGGPSLTS 100
                .         .         .         .         .
    101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 DLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPP 150
                .         .         .         .         .
    151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 AEEELCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVW 200
                .         .
    201 GIEGPIDAAFTRINCQGKTYLFK                           223
        |||||||||||||||||||||||
    201 GIEGPIDAAFTRINCQGKTYLFK                           223
```

DESCRIPTION FOR CLUSTER Z21368

Cluster Z21368 features 7 transcript(s) and 34 segment(s) of interest, the names for which are given in Tables 82 and 83, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 84.

Table 82 - Transcripts of interest

Transcript Name	Sequence ID No.
Z21368_PEA_1_T10	9
Z21368_PEA_1_T11	10
Z21368_PEA_1_T23	11
Z21368_PEA_1_T24	12
Z21368_PEA_1_T5	13
Z21368_PEA_1_T6	14
Z21368_PEA_1_T9	15

Table83 - Segments of interest

Segment Name	Sequence ID No.
Z21368_PEA_1_node_0	1067
Z21368_PEA_1_node_15	1068
Z21368_PEA_1_node_19	1069
Z21368_PEA_1_node_2	1070
Z21368_PEA_1_node_21	1071
Z21368_PEA_1_node_33	1072
Z21368_PEA_1_node_36	1073
Z21368_PEA_1_node_37	1074
Z21368_PEA_1_node_39	1075
Z21368_PEA_1_node_4	1076

Z21368_PEA_1_node_41	1077
Z21368_PEA_1_node_43	1078
Z21368_PEA_1_node_45	1079
Z21368_PEA_1_node_53	1080
Z21368_PEA_1_node_56	1081
Z21368_PEA_1_node_58	1082
Z21368_PEA_1_node_66	1083
Z21368_PEA_1_node_67	1084
Z21368_PEA_1_node_69	1085
Z21368_PEA_1_node_11	1086
Z21368_PEA_1_node_12	1087
Z21368_PEA_1_node_16	1088
Z21368_PEA_1_node_17	1089
Z21368_PEA_1_node_23	1090
Z21368_PEA_1_node_24	1091
Z21368_PEA_1_node_30	1092
Z21368_PEA_1_node_31	1093
Z21368_PEA_1_node_38	1094
Z21368_PEA_1_node_47	1095
Z21368_PEA_1_node_49	1096
Z21368_PEA_1_node_51	1097
Z21368_PEA_1_node_61	1098
Z21368_PEA_1_node_68	1099
Z21368_PEA_1_node_7	1100

Table 84 - Proteins of interest

Protein Name	Sequence ID No.
Z21368_PEA_1_P2	1289
Z21368_PEA_1_P5	1290

Z21368_PEA_1_P15	1291
Z21368_PEA_1_P16	1292
Z21368_PEA_1_P22	1293
Z21368_PEA_1_P23	1294

These sequences are variants of the known protein Extracellular sulfatase Sulf-1 precursor (SwissProt accession identifier SUL1_HUMAN; known also according to the synonyms EC 3.1.6.-; HSulf-1), SEQ ID NO: 1419, referred to herein as the previously known protein.

Protein Extracellular sulfatase Sulf-1 precursor is known or believed to have the following function(s): Exhibits arylsulfatase activity and highly specific endoglucosamine-6-sulfatase activity. It can remove sulfate from the C-6 position of glucosamine within specific subregions of intact heparin. Diminishes HSPG (heparan sulfate proteoglycans) sulfation, inhibits signaling by heparin-dependent growth factors, diminishes proliferation, and facilitates apoptosis in response to exogenous stimulation. The sequence for protein Extracellular sulfatase Sulf-1 precursor is given at the end of the application, as "Extracellular sulfatase Sulf-1 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 85.

Table 85 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
87 – 88	CC->AA: LOSS OF ARYLSULFATASE ACTIVITY AND LOSS OF ABILITY TO MODULATE APOPTOSIS.
49	L -> P
728	K -> R

Protein Extracellular sulfatase Sulf-1 precursor localization is believed to be Endoplasmic reticulum and Golgi stack. Also localized on the cell surface (By similarity).

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: apoptosis; metabolism; heparan sulfate proteoglycan metabolism, which are annotation(s) related to Biological Process; arylsulfatase; hydrolase, which are

annotation(s) related to Molecular Function; and extracellular space; endoplasmic reticulum; Golgi apparatus, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster Z21368 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 13 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 13 and Table 86. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Table 86 - Normal tissue distribution

Name of Tissue	Number
bladder	123
Bone	557
Brain	34
Colon	94
epithelial	56
general	68
head and neck	0
kidney	35
Lung	22
Lymph nodes	0
Breast	52

muscle	31
Ovary	0
pancreas	0
prostate	44
Skin	67
stomach	109
T cells	0
Thyroid	0
Uterus	140

Table 87 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bladder	5.4e-01	6.6e-01	6.4e-01	1.0	8.5e-01	0.7
Bone	4.5e-01	8.2e-01	9.1e-01	0.4	1	0.3
Brain	5.5e-01	7.3e-01	1.5e-01	1.5	5.0e-01	0.9
Colon	1.4e-01	2.8e-01	1.0e-01	2.0	3.0e-01	1.4
epithelial	1.1e-03	1.5e-01	1.2e-07	2.1	1.0e-01	1.1
general	1.4e-05	5.3e-02	1.9e-06	1.6	6.7e-01	0.8
head and neck	2.4e-02	7.1e-02	4.6e-01	2.5	7.5e-01	1.4
kidney	8.9e-01	9.0e-01	1	0.4	1	0.4
Lung	3.5e-01	4.1e-01	7.2e-03	2.6	1.0e-01	1.6
Lymph nodes	7.7e-02	3.1e-01	2.3e-02	8.5	1.9e-01	3.2
Breast	4.0e-01	6.1e-01	5.4e-02	2.3	3.0e-01	1.3
muscle	7.5e-02	3.5e-02	1	1.0	1.7e-01	1.7
Ovary	3.8e-01	4.2e-01	2.2e-01	2.9	3.4e-01	2.2
pancreas	2.2e-02	6.9e-02	1.4e-08	6.5	1.4e-06	4.6
prostate	8.3e-01	8.9e-01	3.1e-01	1.4	5.2e-01	1.1
Skin	6.1e-01	8.1e-01	6.0e-01	1.2	1	0.3
stomach	4.4e-02	5.0e-01	5.0e-01	0.8	9.7e-01	0.4

T cells	5.0e-01	6.7e-01	3.3e-01	3.1	7.2e-01	1.4
Thyroid	3.6e-01	3.6e-01	1	1.1	1	1.1
Uterus	3.5e-01	7.8e-01	4.6e-01	0.9	9.1e-01	0.5

As noted above, cluster Z21368 features 7 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Extracellular sulfatase Sulf-1 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein Z21368_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T5. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P2 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to

MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW
QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR
NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ
FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD
SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP
GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFL
VERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGK
LRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQ
GTPKYKPRFVHTRQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQ
ASSGGNRGRMLADSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYI
DKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKE
AAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWN

corresponding to amino acids 1 - 761 of SUL1_HUMAN, which also corresponds to amino acids 1 - 761 of Z21368_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PHKYSAHGRTRHFESATRTTNGAQKLSRI corresponding to amino acids 762 - 790 of Z21368_PEA_1_P2, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z21368_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PHKYSAHGRTRHFESATRTTNGAQKLSRI in Z21368_PEA_1_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein Z21368_PEA_1_P2 is encoded by the following transcript(s): Z21368_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T5 is shown in bold; this coding portion starts at position 529 and ends at position 2898.

Variant protein Z21368_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T9. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the

relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P5 and Q7Z2W2 (SEQ ID NO:1697):

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVEL corresponding to amino acids 1 - 57 of Q7Z2W2, which also corresponds to amino acids 1 - 57 of Z21368_PEA_1_P5, second bridging amino acid sequence comprising A, and a third amino acid sequence being at least 90 % homologous to FFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITN ESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNM DKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYT ADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDT PPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHL PKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLY ARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFE GEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPT TVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKR RKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKER KEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNE THNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCN PRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 139 - 871 of Q7Z2W2, which also corresponds to amino acids 59 - 791 of Z21368_PEA_1_P5, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for an edge portion of Z21368_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least three amino acids comprise LAF, the sequence having a structure as follows (numbering according to Z21368_PEA_1_P5): a

sequence starting from any of amino acid numbers 57-x to 57; and ending at any of amino acid numbers 59 + ((n-2) - x), in which x varies from 0 to n-2.

Comparison report between Z21368_PEA_1_P5 and AAH12997 (SEQ ID NO:1698):

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELAFF GKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNES INYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNMDK HWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYTAD HGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPP DVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLP KYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYA RGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFEGE IYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTV RVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRK PEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKE KRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETH NFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLME corresponding to amino acids 1 - 751 of Z21368_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to LRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 1 - 40 of AAH12997, which also corresponds to amino acids 752 - 791 of Z21368_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of Z21368_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELAFF

GKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNES
INYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPNMDK
HWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYTAD
HGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPP
DVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLP
KYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYA
RGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFEGE
IYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTV
RVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRK
PEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKE
KRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETH
NFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLME of
Z21368_PEA_1_P5.

Comparison report between Z21368_PEA_1_P5 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVEL
corresponding to amino acids 1 - 57 of SUL1_HUMAN, which also corresponds to amino acids 1 - 57 of Z21368_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to
AFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLIT
NESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYNYAPN
MDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYII
YTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGL
DTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSN
HLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRN
LYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVE
FEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGP
PTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHL
KRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENNRRRK
KERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRT

VNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYK QCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG corresponding to amino acids 138 - 871 of SUL1_HUMAN, which also corresponds to amino acids 58 - 791 of Z21368_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of Z21368_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise LA, having a structure as follows: a sequence starting from any of amino acid numbers 57-x to 57; and ending at any of amino acid numbers 58 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z21368_PEA_1_P5 is encoded by the following transcript(s): Z21368_PEA_1_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T9 is shown in bold; this coding portion starts at position 556 and ends at position 2928.

Variant protein Z21368_PEA_1_P15 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T23. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the

relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P15 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P15, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFL VERG corresponding to amino acids 1 - 416 of SUL1_HUMAN, which also corresponds to amino acids 1 - 416 of Z21368_PEA_1_P15.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z21368_PEA_1_P15 is encoded by the following transcript(s): Z21368_PEA_1_T23, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T23 is shown in bold; this coding portion starts at position 691 and ends at position 1938.

Variant protein Z21368_PEA_1_P16 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T24. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously

published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P16 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P16, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR NGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQ FSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDD SVERLYNMLVETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEP GSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNR corresponding to amino acids 1 - 397 of SUL1_HUMAN, which also corresponds to amino acids 1 - 397 of Z21368_PEA_1_P16, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence CVIVPPLSQPQIH corresponding to amino acids 398 - 410 of Z21368_PEA_1_P16, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z21368_PEA_1_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence CVIVPPLSQPQIH in Z21368_PEA_1_P16.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z21368_PEA_1_P16 is encoded by the following transcript(s): Z21368_PEA_1_T24, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T24 is shown in bold; this coding portion starts at position 691 and ends at position 1920.

Variant protein Z21368_PEA_1_P22 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T10. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P22 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P22, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCR NGIKEKHGFDYAK corresponding to amino acids 1 - 188 of SUL1_HUMAN, which also corresponds to amino acids 1 - 188 of Z21368_PEA_1_P22, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ARYDGDQPRCAPRPRGLSPTVF corresponding to amino acids 189 - 210 of Z21368_PEA_1_P22, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z21368_PEA_1_P22, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ARYDGDQPRCAPRPRGLSPTVF in Z21368_PEA_1_P22.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z21368_PEA_1_P22 is encoded by the following transcript(s): Z21368_PEA_1_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T10 is shown in bold; this coding portion starts at position 691 and ends at position 1320.

Variant protein Z21368_PEA_1_P23 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z21368_PEA_1_T11. An alignment is given to the known protein (Extracellular sulfatase Sulf-1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z21368_PEA_1_P23 and Q7Z2W2:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P23, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRT corresponding to amino acids 1 - 137 of Q7Z2W2, which also corresponds to amino acids 1 - 137 of Z21368_PEA_1_P23, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

GLLHRLNH corresponding to amino acids 138 - 145 of Z21368_PEA_1_P23, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z21368_PEA_1_P23, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GLLHRLNH in Z21368_PEA_1_P23.

Comparison report between Z21368_PEA_1_P23 and SUL1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z21368_PEA_1_P23, comprising a first amino acid sequence being at least 90 % homologous to MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSW QAMHEPRTFAVYLNNTGYRT corresponding to amino acids 1 - 137 of SUL1_HUMAN, which also corresponds to amino acids 1 - 137 of Z21368_PEA_1_P23, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GLLHRLNH corresponding to amino acids 138 - 145 of Z21368_PEA_1_P23, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail ofZ21368_PEA_1_P23, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GLLHRLNH in Z21368_PEA_1_P23.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z21368_PEA_1_P23 is encoded by the following transcript(s): Z21368_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z21368_PEA_1_T11 is shown in bold; this coding portion starts at position 691 and ends at position 1125.

As noted above, cluster Z21368 features 34 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster Z21368_PEA_1_node_0 according to the present invention is supported by 8 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T9. Table 88 below describes the starting and ending position of this segment on each transcript.

Table 88 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T9	1	327

Segment cluster Z21368_PEA_1_node_15 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 89 below describes the starting and ending position of this segment on each transcript.

Table 89 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	631	807
Z21368_PEA_1_T11	631	807

Z21368_PEA_1_T23	631	807
Z21368_PEA_1_T24	631	807
Z21368_PEA_1_T5	469	645
Z21368_PEA_1_T6	469	645
Z21368_PEA_1_T9	496	672

Segment cluster Z21368_PEA_1_node_19 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5 and Z21368_PEA_1_T6. Table 90 below describes the starting and ending position of this segment on each transcript.

Table 90 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	863	1102
Z21368_PEA_1_T11	863	1102
Z21368_PEA_1_T23	863	1102
Z21368_PEA_1_T24	863	1102
Z21368_PEA_1_T5	701	940
Z21368_PEA_1_T6	701	940

Segment cluster Z21368_PEA_1_node_2 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5 and Z21368_PEA_1_T6. Table 91 below describes the starting and ending position of this segment on each transcript.

Table 91 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

Z21368_PEA_1_T10	1	300
Z21368_PEA_1_T11	1	300
Z21368_PEA_1_T23	1	300
Z21368_PEA_1_T24	1	300
Z21368_PEA_1_T5	1	300
Z21368_PEA_1_T6	1	300

Segment cluster Z21368_PEA_1_node_21 according to the present invention is supported by 37 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 92 below describes the starting and ending position of this segment on each transcript.

Table 92 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1103	1254
Z21368_PEA_1_T23	1103	1254
Z21368_PEA_1_T24	1103	1254
Z21368_PEA_1_T5	941	1092
Z21368_PEA_1_T6	941	1092
Z21368_PEA_1_T9	728	879

Segment cluster Z21368_PEA_1_node_33 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 93 below describes the starting and ending position of this segment on each transcript.

Table 93 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1502	1677
Z21368_PEA_1_T11	1424	1599
Z21368_PEA_1_T23	1576	1751
Z21368_PEA_1_T24	1576	1751
Z21368_PEA_1_T5	1414	1589
Z21368_PEA_1_T6	1414	1589
Z21368_PEA_1_T9	1201	1376

Segment cluster Z21368_PEA_1_node_36 according to the present invention is supported by 44 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 94 below describes the starting and ending position of this segment on each transcript.

Table 94 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1678	1806
Z21368_PEA_1_T11	1600	1728
Z21368_PEA_1_T23	1752	1880
Z21368_PEA_1_T24	1752	1880
Z21368_PEA_1_T5	1590	1718
Z21368_PEA_1_T6	1590	1718
Z21368_PEA_1_T9	1377	1505

Segment cluster Z21368_PEA_1_node_37 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): Z21368_PEA_1_T24. Table 95 below describes the starting and ending position of this segment on each transcript.

Table 95 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T24	1881	2159

Segment cluster Z21368_PEA_1_node_39 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T23 and Z21368_PEA_1_T24. Table 96 below describes the starting and ending position of this segment on each transcript.

Table 96 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T23	1938	2790
Z21368_PEA_1_T24	2217	3069

Segment cluster Z21368_PEA_1_node_4 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23 and Z21368_PEA_1_T24. Table 97 below describes the starting and ending position of this segment on each transcript.

Table 97 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	301	462
Z21368_PEA_1_T11	301	462
Z21368_PEA_1_T23	301	462
Z21368_PEA_1_T24	301	462

Segment cluster Z21368_PEA_1_node_41 according to the present invention is supported by 49 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 98 below describes the starting and ending position of this segment on each transcript.

Table 98 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1864	1993
Z21368_PEA_1_T11	1786	1915
Z21368_PEA_1_T5	1776	1905
Z21368_PEA_1_T6	1776	1905
Z21368_PEA_1_T9	1563	1692

Segment cluster Z21368_PEA_1_node_43 according to the present invention is supported by 52 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 99 below describes the starting and ending position of this segment on each transcript.

Table 99 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1994	2210
Z21368_PEA_1_T11	1916	2132
Z21368_PEA_1_T5	1906	2122
Z21368_PEA_1_T6	1906	2122
Z21368_PEA_1_T9	1693	1909

Segment cluster Z21368_PEA_1_node_45 according to the present invention is supported by 64 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 100 below describes the starting and ending position of this segment on each transcript.

Table 100 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2211	2466
Z21368_PEA_1_T11	2133	2388
Z21368_PEA_1_T5	2123	2378
Z21368_PEA_1_T6	2123	2378
Z21368_PEA_1_T9	1910	2165

Segment cluster Z21368_PEA_1_node_53 according to the present invention is supported by 60 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 101 below describes the starting and ending position of this segment on each transcript.

Table 102 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2725	2900
Z21368_PEA_1_T11	2647	2822
Z21368_PEA_1_T5	2637	2812
Z21368_PEA_1_T6	2637	2812
Z21368_PEA_1_T9	2424	2599

Segment cluster Z21368_PEA_1_node_56 according to the present invention is supported by 50 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11 and Z21368_PEA_1_T9. Table 102 below describes the starting and ending position of this segment on each transcript.

Table 102 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2901	3043
Z21368_PEA_1_T11	2823	2965
Z21368_PEA_1_T9	2600	2742

Segment cluster Z21368_PEA_1_node_58 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 103 below describes the starting and ending position of this segment on each transcript.

Table 103 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	3044	3167
Z21368_PEA_1_T11	2966	3089
Z21368_PEA_1_T5	2813	2936
Z21368_PEA_1_T6	2813	2936
Z21368_PEA_1_T9	2743	2866

Segment cluster Z21368_PEA_1_node_66 according to the present invention is supported by 142 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11,

Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 104 below describes the starting and ending position of this segment on each transcript.

Table 104 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	3202	3789
Z21368_PEA_1_T11	3124	3711
Z21368_PEA_1_T5	2971	3558
Z21368_PEA_1_T6	2971	3558
Z21368_PEA_1_T9	2901	3488

Segment cluster Z21368_PEA_1_node_67 according to the present invention is supported by 181 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 105 below describes the starting and ending position of this segment on each transcript.

Table 105 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	3790	4374
Z21368_PEA_1_T11	3712	4296
Z21368_PEA_1_T5	3559	4143
Z21368_PEA_1_T6	3559	4143
Z21368_PEA_1_T9	3489	4073

Segment cluster Z21368_PEA_1_node_69 according to the present invention is supported by 150 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11,

Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 106 below describes the starting and ending position of this segment on each transcript.

Table 107 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	4428	4755
Z21368_PEA_1_T11	4350	4677
Z21368_PEA_1_T5	4197	5384
Z21368_PEA_1_T6	4197	4524
Z21368_PEA_1_T9	4127	4454

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster Z21368_PEA_1_node_11 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 107 below describes the starting and ending position of this segment on each transcript.

Table 107 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	558	602
Z21368_PEA_1_T11	558	602
Z21368_PEA_1_T23	558	602
Z21368_PEA_1_T24	558	602
Z21368_PEA_1_T5	396	440
Z21368_PEA_1_T6	396	440
Z21368_PEA_1_T9	423	467

Segment cluster Z21368_PEA_1_node_12 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 108 below describes the starting and ending position of this segment on each transcript.

Table 108 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	603	630
Z21368_PEA_1_T11	603	630
Z21368_PEA_1_T23	603	630
Z21368_PEA_1_T24	603	630
Z21368_PEA_1_T5	441	468
Z21368_PEA_1_T6	441	468
Z21368_PEA_1_T9	468	495

Segment cluster Z21368_PEA_1_node_16 according to the present invention can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 109 below describes the starting and ending position of this segment on each transcript.

Table 109 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	808	822
Z21368_PEA_1_T11	808	822
Z21368_PEA_1_T23	808	822
Z21368_PEA_1_T24	808	822

Z21368_PEA_1_T5	646	660
Z21368_PEA_1_T6	646	660
Z21368_PEA_1_T9	673	687

Segment cluster Z21368_PEA_1_node_17 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 110 below describes the starting and ending position of this segment on each transcript.

Table 110 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	823	862
Z21368_PEA_1_T11	823	862
Z21368_PEA_1_T23	823	862
Z21368_PEA_1_T24	823	862
Z21368_PEA_1_T5	661	700
Z21368_PEA_1_T6	661	700
Z21368_PEA_1_T9	688	727

Segment cluster Z21368_PEA_1_node_23 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 111 below describes the starting and ending position of this segment on each transcript.

Table 111 Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

Z21368_PEA_1_T11	1103	1176
Z21368_PEA_1_T23	1255	1328
Z21368_PEA_1_T24	1255	1328
Z21368_PEA_1_T5	1093	1166
Z21368_PEA_1_T6	1093	1166
Z21368_PEA_1_T9	880	953

Segment cluster Z21368_PEA_1_node_24 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 112 below describes the starting and ending position of this segment on each transcript.

Table 112 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1255	1350
Z21368_PEA_1_T11	1177	1272
Z21368_PEA_1_T23	1329	1424
Z21368_PEA_1_T24	1329	1424
Z21368_PEA_1_T5	1167	1262
Z21368_PEA_1_T6	1167	1262
Z21368_PEA_1_T9	954	1049

Segment cluster Z21368_PEA_1_node_30 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and

Z21368_PEA_1_T9. Table 113 below describes the starting and ending position of this segment on each transcript.

Table 113 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1351	1409
Z21368_PEA_1_T11	1273	1331
Z21368_PEA_1_T23	1425	1483
Z21368_PEA_1_T24	1425	1483
Z21368_PEA_1_T5	1263	1321
Z21368_PEA_1_T6	1263	1321
Z21368_PEA_1_T9	1050	1108

Segment cluster Z21368_PEA_1_node_31 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 114 below describes the starting and ending position of this segment on each transcript.

Table 114 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1410	1501
Z21368_PEA_1_T11	1332	1423
Z21368_PEA_1_T23	1484	1575
Z21368_PEA_1_T24	1484	1575
Z21368_PEA_1_T5	1322	1413
Z21368_PEA_1_T6	1322	1413
Z21368_PEA_1_T9	1109	1200

Segment cluster Z21368_PEA_1_node_38 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 115 below describes the starting and ending position of this segment on each transcript.

Table 115 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	1807	1863
Z21368_PEA_1_T11	1729	1785
Z21368_PEA_1_T23	1881	1937
Z21368_PEA_1_T24	2160	2216
Z21368_PEA_1_T5	1719	1775
Z21368_PEA_1_T6	1719	1775
Z21368_PEA_1_T9	1506	1562

Segment cluster Z21368_PEA_1_node_47 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 116 below describes the starting and ending position of this segment on each transcript.

Table 116 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2467	2563
Z21368_PEA_1_T11	2389	2485
Z21368_PEA_1_T5	2379	2475
Z21368_PEA_1_T6	2379	2475

Z21368_PEA_1_T9	2166	2262

Segment cluster Z21368_PEA_1_node_49 according to the present invention is supported by 57 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 117 below describes the starting and ending position of this segment on each transcript.

Table 117 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2564	2658
Z21368_PEA_1_T11	2486	2580
Z21368_PEA_1_T5	2476	2570
Z21368_PEA_1_T6	2476	2570
Z21368_PEA_1_T9	2263	2357

Segment cluster Z21368_PEA_1_node_51 according to the present invention is supported by 46 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 118 below describes the starting and ending position of this segment on each transcript.

Table 118 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	2659	2724
Z21368_PEA_1_T11	2581	2646
Z21368_PEA_1_T5	2571	2636
Z21368_PEA_1_T6	2571	2636
Z21368_PEA_1_T9	2358	2423

Segment cluster Z21368_PEA_1_node_61 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 119 below describes the starting and ending position of this segment on each transcript.

Table119 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	3168	3201
Z21368_PEA_1_T11	3090	3123
Z21368_PEA_1_T5	2937	2970
Z21368_PEA_1_T6	2937	2970
Z21368_PEA_1_T9	2867	2900

Segment cluster Z21368_PEA_1_node_68 according to the present invention is supported by 87 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 120 below describes the starting and ending position of this segment on each transcript.

Table 120 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	4375	4427
Z21368_PEA_1_T11	4297	4349
Z21368_PEA_1_T5	4144	4196
Z21368_PEA_1_T6	4144	4196
Z21368_PEA_1_T9	4074	4126

Segment cluster Z21368_PEA_1_node_7 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z21368_PEA_1_T10, Z21368_PEA_1_T11, Z21368_PEA_1_T23, Z21368_PEA_1_T24, Z21368_PEA_1_T5, Z21368_PEA_1_T6 and Z21368_PEA_1_T9. Table 121 below describes the starting and ending position of this segment on each transcript.

Table 121 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z21368_PEA_1_T10	463	557
Z21368_PEA_1_T11	463	557
Z21368_PEA_1_T23	463	557
Z21368_PEA_1_T24	463	557
Z21368_PEA_1_T5	301	395
Z21368_PEA_1_T6	301	395
Z21368_PEA_1_T9	328	422

Overexpression of at least a portion of this cluster was determined according to oligonucleotides and one or more chips. The results were as follows: Oligonucleotide Z21368_0_0_61857 was on the TAA chip and was found to be overexpressed in Lung cancer (general), in Lung adenocarcinoma, and in Lung squamous cell cancer.

Variant protein alignment to the previously known protein:

Sequence name: /tmp/5ER3vIMKE2/9L0Y7lDlTQ:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P2 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 7664.00

Escore: 0

Matching length: 761 Total length: 761

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
                .         .         .         .         .
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
                .         .         .         .         .
    201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
```

```
                                                                    
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
                                                                    
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
                                                                    
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
                                                                    
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
                                                                    
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
                                                                    
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
                                                                    
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
                                                                    
401 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 450
                                                                    
401 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 450
                                                                    
451 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 500
                                                                    
451 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 500
                                                                    
501 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 550
                                                                    
501 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 550
                                                                    
551 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 600
                                                                    
551 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 600
                                                                    
```

```
601 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 650
          .         .         .         .         .
651 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 700
          .         .         .         .         .
701 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 750
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 750
          .
751 NNHWQTAPFWN                                       761
    ||||||||||||
751 NNHWQTAPFWN                                       761
```

Sequence name: /tmp/tt3yfXIUKV/YxSTFWr66h:Q7Z2W2

Sequence documentation:

Alignment of: Z21368_PEA_1_P5 x Q7Z2W2 ..

Alignment segment 1/1:

Quality: 7869.00

Escore: 0

305

Matching length: 791 Total length: 871

Matching Percent Similarity: 99.87 Matching Percent Identity: 99.87

Total Percent Similarity: 90.70 Total Percent Identity: 90.70

Gaps: 1

Alignment:

```
                .         .         .         .         .
  1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
 51 DDQDVELA.......................................... 58
    |||||||
 51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
 59 ....................................FFGKYLNEYNGS 70
                                        ||||||||||||
101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTVFFGKYLNEYNGS 150
                .         .         .         .         .
 71 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
                .         .         .         .         .
121 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 170
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
                .         .         .         .         .
171 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 220
```

```
                                                                    
                                                                    
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
                                                                    
221 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 270
                                                                    
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
                                                                    
271 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 320
                                                                    
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
                                                                    
321 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 370
                                                                    
401 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 450
                                                                    
371 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 420
                                                                    
451 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 500
                                                                    
421 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 470
                                                                    
501 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 550
                                                                    
471 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 520
                                                                    
551 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 600
                                                                    
521 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 570
                                                                    
601 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 650
                                                                    
```

```
571 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 620
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
651 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 700
              .         .         .         .         .
621 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 670
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
701 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 750
              .         .         .         .         .
671 NNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHNFLFCEFATGFLEY 720
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
751 NNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHNFLFCEFATGFLEY 800
              .         .         .         .         .
721 FDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDV 770
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
801 FDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDV 850
              .         .
771 GNKDGGSYDLHRGQLWDGWEG                               791
    |||||||||||||||||||||
851 GNKDGGSYDLHRGQLWDGWEG                               871
```

Sequence name: /tmp/tt3yfXIUKV/YxSTFWr66h:AAH12997

Sequence documentation:

Alignment of: Z21368_PEA_1_P5 x AAH12997 ..

Alignment segment 1/1:

Quality: 420.00

Escore: 0

Matching length: 40 Total length: 40

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .
     752 LRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG  791
         ||||||||||||||||||||||||||||||||||||||||
       1 LRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG   40
```

Sequence name: /tmp/tt3yfXIUKV/YxSTFWr66h:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P5 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 7878.00

Escore: 0

Matching length: 791 Total length: 871

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 90.82 Total Percent Identity: 90.82

Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
       51 DDQDVEL..................................................  57
          |||||||
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
       58 .....................................AFFGKYLNEYNGS  70
                                               |||||||||||||
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
                .         .         .         .         .
       71 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
                .         .         .         .         .
      121 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 170
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
```

```
                .         .         .         .         .
171 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 220
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
                .         .         .         .         .
221 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 270
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
                .         .         .         .         .
271 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 320
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
                .         .         .         .         .
321 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 370
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 NKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARY 450
                .         .         .         .         .
371 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 420
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 QTACEQPGQKWQCIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDK 500
                .         .         .         .         .
421 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 470
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 DKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEF 550
                .         .         .         .         .
471 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 520
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 EGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA 600
                .         .         .         .         .
521 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 570
    ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
601 DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEI 650
                .         .         .         .         .
571 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 620
    |||||||||||||||||||||||||||||||||||||||||||||||||||
651 EALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLH 700
                .         .         .         .         .
621 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 670
    |||||||||||||||||||||||||||||||||||||||||||||||||||
701 PFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHD 750
                .         .         .         .         .
671 NNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHNFLFCEFATGFLEY 720
    |||||||||||||||||||||||||||||||||||||||||||||||||||
751 NNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHNFLFCEFATGFLEY 800
                .         .         .         .         .
721 FDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDV 770
    |||||||||||||||||||||||||||||||||||||||||||||||||||
801 FDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDV 850
                .         .
771 GNKDGGSYDLHRGQLWDGWEG                               791
    |||||||||||||||||||
851 GNKDGGSYDLHRGQLWDGWEG                               871
```

Sequence name: /tmp/AVAZGWHuF0/RzHFOnHIsT:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P15 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 4174.00

Escore: 0

Matching length: 416 Total length: 416

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
                .         .         .         .         .
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
```

```
                .         .         .         .         .
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
                .         .         .         .         .
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
                .         .         .         .         .
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
    |||||||||||||||||||||||||||||||||||||||||||||||||||
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
                .         .         .         .         .
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
    |||||||||||||||||||||||||||||||||||||||||||||||||||
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRT 400
                .
401 NKKAKIWRDTFLVERG                                   416
    ||||||||||||||||
401 NKKAKIWRDTFLVERG                                   416
```

Sequence name: /tmp/JhwgRdKqmt/kqSmjxkWWk:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P16 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 3985.00
Escore: 0
Matching length: 397 Total length: 397
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
                .         .         .         .         .
      151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESI 200
                .         .         .         .         .
```

315

```
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 NYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPNASQHITPSYN 250
              .         .         .         .         .
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 YAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML 300
              .         .         .         .         .
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVE 350
              .         .         .         .
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNR    397
    |||||||||||||||||||||||||||||||||||||||||||||||||
351 PGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNR    397
```

Sequence name: /tmp/GPlnIw3BOg/zXFdxqG4ow:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P22 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 1897.00

Escore: 0

316

Matching length: 188 Total length: 188

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .         .         .
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGS 150
                .         .         .
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAK             188
        ||||||||||||||||||||||||||||||||||||
    151 YIPPGWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAK             188
```

Sequence name: /tmp/oji5Fs74fB/8xeB9KrGjp:Q7Z2W2

Sequence documentation:

Alignment of: Z21368_PEA_1_P23 x Q7Z2W2 ..

Alignment segment 1/1:

Quality: 1368.00

Escore: 0.000511

Matching length: 137 Total length: 137

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRT             137
          |||||||||||||||||||||||||||||||||||
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRT             137
```

Sequence name: /tmp/oji5Fs74fB/8xeB9KrGjp:SUL1_HUMAN

Sequence documentation:

Alignment of: Z21368_PEA_1_P23 x SUL1_HUMAN ..

Alignment segment 1/1:

Quality: 1368.00

Escore: 0.000511

Matching length: 137 Total length: 137

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLT 50
                .         .         .         .         .
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 DDQDVELGSLQVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYV 100
                .         .         .
      101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRT              137
```

```
                                319
                                 |
                                 |||||||||||||||||||||||||||||||||||||
101 HNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRT                          137
```

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1Z21368 transcripts which are detectable by amplicon as depicted in sequence name Z21368junc17-21 in normal and cancerous lung tissues

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by or according to junc17-21 segment, Z21368junc17-21 amplicon (SEQ ID NO: 1642) and Z21368junc17-21F (SEQ ID NO: 1640) Z21368junc17-21R (SEQ ID NO: 1641) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 14 is a histogram showing over expression of the above-indicated SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained. As is evident from Figure 14, the expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by the above amplicon in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 10 out of 15 adenocarcinoma samples, 7 out of 16 squamous cell carcinoma samples, 0 out of 4 large cell carcinoma samples and in 0 out of 8 small cells carcinoma samples.

Threshold of 5 fold over-expression was found to differentiate between cancer and normal samples with P value of 3.56E-04 in adenocarcinoma, 9.66E-03 in squamous cell carcinomas checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: *Z21368junc17-21F* forward primer; and *Z21368junc17-21 R* reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: *Z21368junc17-21*.

Forward primer (SEQ ID NO: 1640): GGACGGATACAGCAGGAACG

Reverse amplicon (SEQ ID NO: 1641): TATTTTCCAAAAAAGGCCAGCTC

Amplicon (SEQ ID NO: 1642):

GGACGGATACAGCAGGAACGAAAAAAACATCCGACCCAACATTATTCTTGTGCTTAC CGATGATCAAGATGTGGAGCTGGCCTTTTTTGGAAAATA

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368 junc17-21 in different normal tissues

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by or according to Z21368 junc17-21 amplicon (SEQ ID NO: 1642) and Z21368 junc17-21F (SEQ ID NO: 1640) and Z21368 junc17-21R (SEQ ID NO: 1641) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession

No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the breast samples (Sample Nos. 33-35 Table 3, "Tissue samples in normal panel", above), to obtain a value of relative expression of each sample relative to median of the breast samples.

Forward primer (SEQ ID NO: 1640): GGACGGATACAGCAGGAACG
Reverse amplicon (SEQ ID NO: 1641): TATTTTCCAAAAAAGGCCAGCTC
Amplicon (SEQ ID NO: 1642):
GGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTAC
CGATGATCAAGATGTGGAGCTGGCCTTTTTTGGAAAATA

The results are shown in Figure 15, demonstrating the expression of Extracellular sulfatase Sulf-1Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368 junc17-21, in different normal tissues.

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 Z21368 transcripts which are detectable by amplicon as depicted in sequence name Z21368seg39 in normal and cancerous lung tissues

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by or according to seg39, Z21368seg39 amplicon (SEQ ID NO: 1645) and primers Z21368seg39F (SEQ ID NO: 1643) and Z21368seg39R (SEQ ID NO: 1644) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the

geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 16 is a histogram showing over expression of the above-indicated SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.

As is evident from Figure 16, the expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by the above amplicon in cancer samples was higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 8 out of 15 adenocarcinoma samples, 5 out of 16 squamous cell carcinoma samples and 1 out of 4 large cell carcinoma samples .

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by the above amplicon in lung cancer samples versus the normal tissue samples was determined by T test as 2.17E-04 in adenocarcinoma, 9.94E-03 in squamous cell carcinoma and 2.17E-01 in large cell carcinoma.

Threshold of 5fold overexpression was found to differentiate between cancer and normal samples with P value of 1.74E-02 in adenocarcinoma, 1.58E-01 in squamous cell carcinoma and 4.33E-01 in large cell carcinoma as checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Z21368seg39F forward primer; and Z21368seg39R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Z21368seg39.

Primers:

Forward primer Z21368seg39F (SEQ ID NO: 1643): GTTGCATTTCTCAGTGCTGGTTT

Reverse primer Z21368seg39R (SEQ ID NO: 1644): AGGGTGCCGGGTGAGG

Amplicon Z21368seg39 (SEQ ID NO: 1645):

GTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTCTACCATC CTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCT

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1Z21368 transcripts which are detectable by amplicon as depicted in sequence name Z21368seg39 in different normal tissues

Expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1 transcripts detectable by or according to *Z21368seg39* amplicon (SEQ ID NO: 1645) and *Z21368seg39F* (SEQ ID NO: 1643) *Z21368seg39R* (SEQ ID NO: 1644) was measured by real time PCR. In parallel the expression of four housekeeping genes –[RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the breast samples (Sample Nos. 33-35 Table 3, above), to obtain a value of relative expression of each sample relative to median of the breast samples.

Forward primer Z21368seg39F (SEQ ID NO: 1643): GTTGCATTTCTCAGTGCTGGTTT
Reverse primer Z21368seg39R (SEQ ID NO: 1644): AGGGTGCCGGGTGAGG
Amplicon Z21368seg39 (SEQ ID NO: 1645):
GTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTCTACCATC
CTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCT

The results are demonstrated in Figure 17, showing expression of SUL1_HUMAN - Extracellular sulfatase Sulf-1, Z21368 transcripts, which are detectable by amplicon as depicted in sequence name Z21368seg39, in different normal tissues.

PBGD-amplicon, SEQ ID NO:334HPRT1-amplicon, SEQ ID NO:1297Ubiquitin-amplicon, SEQ ID NO:328SDHA-amplicon, SEQ ID NO:331PBGD-amplicon, SEQ ID NO:334HPRT1-amplicon, SEQ ID NO:1297Ubiquitin-amplicon, SEQ ID NO:328SDHA-amplicon, SEQ ID NO:331RPL19 amplicon, SEQ ID NO:1630TATA amplicon, SEQ ID NO:1633Ubiquitin-amplicon, SEQ ID NO:328SDHA-amplicon, SEQ ID NO:331

DESCRIPTION FOR CLUSTER HUMGRP5E

Cluster HUMGRP5E features 2 transcript(s) and 5 segment(s) of interest, the names for which are given in Tables 160 and 161, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 162.

Table 160 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMGRP5E_T4	20

HUMGRP5E_T5	21

Table 161 - Segments of interest

Segment Name	Sequence ID No.
HUMGRP5E_node_0	335
HUMGRP5E_node_2	336
HUMGRP5E_node_8	337
HUMGRP5E_node_3	338
HUMGRP5E_node_7	339

Table 162 - Proteins of interest

Protein Name	Sequence ID No.
HUMGRP5E_P4	1299
HUMGRP5E_P5	1300

These sequences are variants of the known protein Gastrin-releasing peptide precursor (SwissProt accession identifier GRP_HUMAN; known also according to the synonyms GRP; GRP-10), SEQ ID NO: 1421, referred to herein as the previously known protein.

Gastrin-releasing peptide is known or believed to have the following function(s): stimulates gastrin release as well as other gastrointestinal hormones. The sequence for protein Gastrin-releasing peptide precursor is given at the end of the application, as "Gastrin-releasing peptide precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 163.

Table163 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
4	S -> R

Protein Gastrin-releasing peptide localization is believed to be Secreted.

The previously known protein also has the following indication(s) and/or potential therapeutic use(s): Diabetes, Type II. It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Bombesin antagonist; Insulinotropin agonist. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Anorectic/Antiobesity; Releasing hormone; Anticancer; Respiratory; Antidiabetic.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: signal transduction; neuropeptide signaling pathway, which are annotation(s) related to Biological Process; growth factor, which are annotation(s) related to Molecular Function; and secreted, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster HUMGRP5E features 2 transcript(s), which were listed in Table 160 above. These transcript(s) encode for protein(s) which are variant(s) of protein Gastrin-releasing peptide precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HUMGRP5E_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMGRP5E_T4. An alignment is given to the known protein (Gastrin-releasing peptide precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMGRP5E_P4 and GRP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMGRP5E_P4, comprising a first amino acid sequence being at least 90 % homologous to MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTG ESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSED SSNFKDVGSKGK corresponding to amino acids 1 - 127 of GRP_HUMAN, which also corresponds to amino acids 1 - 127 of HUMGRP5E_P4, and a second amino acid sequence being at least 90 % homologous to GSQREGRNPQLNQQ corresponding to amino acids 135 - 148 of GRP_HUMAN, which also corresponds to amino acids 128 - 141 of HUMGRP5E_P4, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HUMGRP5E_P4, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KG, having a structure as follows: a sequence starting from any of amino acid numbers 127-x to 127; and ending at any of amino acid numbers 128 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein HUMGRP5E_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 164, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMGRP5E_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 164 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	S -> R	Yes

Variant protein HUMGRP5E_P4 is encoded by the following transcript(s): HUMGRP5E_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMGRP5E_T4 is shown in bold; this coding portion starts at position 622 and ends at position 1044. The transcript also has the following SNPs as listed in Table 165 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMGRP5E_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 165 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
541	-> T	No
542	G -> T	No
631	A -> C	Yes
672	G -> A	Yes
1340	C ->	No
1340	C -> A	No
1341	A ->	No
1341	A -> G	No

Variant protein HUMGRP5E_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMGRP5E_T5. An alignment is given to the known protein (Gastrin-releasing peptide precursor) at the end of the application. One or more alignments to one or more previously published protein sequences

are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMGRP5E_P5 and GRP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMGRP5E_P5, comprising a first amino acid sequence being at least 90 % homologous to MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTG ESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSED SSNFKDVGSKGK corresponding to amino acids 1 - 127 of GRP_HUMAN, which also corresponds to amino acids 1 - 127 of HUMGRP5E_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DSLLQVLNVKEGTPS corresponding to amino acids 128 - 142 of HUMGRP5E_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMGRP5E_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DSLLQVLNVKEGTPS in HUMGRP5E_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMGRP5E_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 166, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMGRP5E_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 166 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	S -> R	Yes

Variant protein HUMGRP5E_P5 is encoded by the following transcript(s): HUMGRP5E_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMGRP5E_T5 is shown in bold; this coding portion starts at position 622 and ends at position 1047. The transcript also has the following SNPs as listed in Table 167 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMGRP5E_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 167 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
541	-> T	No
542	G -> T	No
631	A -> C	Yes
672	G -> A	Yes
1354	C ->	No
1354	C -> A	No
1355	A ->	No
1355	A -> G	No

As noted above, cluster HUMGRP5E features 5 segment(s), which were listed in Table 161 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMGRP5E_node_0 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMGRP5E_T4 and HUMGRP5E_T5. Table 168 below describes the starting and ending position of this segment on each transcript.

Table 168 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMGRP5E_T4	1	760
HUMGRP5E_T5	1	760

Segment cluster HUMGRP5E_node_2 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMGRP5E_T4 and HUMGRP5E_T5. Table 169 below describes the starting and ending position of this segment on each transcript.

Table 169 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMGRP5E_T4	761	984
HUMGRP5E_T5	761	984

Segment cluster HUMGRP5E_node_8 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMGRP5E_T4 and HUMGRP5E_T5. Table 170 below describes the starting and ending position of this segment on each transcript.

Table 170 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMGRP5E_T4	1004	1362
HUMGRP5E_T5	1018	1376

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMGRP5E_node_3 according to the present invention can be found in the following transcript(s): HUMGRP5E_T4 and HUMGRP5E_T5. Table 171 below describes the starting and ending position of this segment on each transcript.

Table 171 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMGRP5E_T4	985	1003
HUMGRP5E_T5	985	1003

Segment cluster HUMGRP5E_node_7 according to the present invention can be found in the following transcript(s): HUMGRP5E_T5. Table 172 below describes the starting and ending position of this segment on each transcript.

Table 172 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMGRP5E_T5	1004	1017

Microarray (chip) data is also available for this gene as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (with regard to lung cancer), shown in Table 173.

Table 173 - Oligonucleotides related to this gene

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMGRP5E_0_0_16630	Lung cancer	Lung
HUMGRP5E_0_2_0	Lung cancer	Lung

Variant protein alignment to the previously known protein:

Sequence name: /tmp/412zs2mwyT/B0wjOUAX0d:GRP_HUMAN

Sequence documentation:

Alignment of: HUMGRP5E_P4 x GRP_HUMAN ..

Alignment segment 1/1:

Quality: 1291.00

Escore: 0

Matching length: 141 Total length: 148

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 95.27 Total Percent Identity: 95.27

Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLM 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLM 50
                .         .         .         .         .
     51 GKKSTGESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQ 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GKKSTGESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQ 100
                .         .         .         .
```

```
101 PKALGNQQPSWDSEDSSNFKDVGSKGK.......GSQREGRNPQLNQQ   141
    ||||||||||||||||||||||||||||       ||||||||||||||
101 PKALGNQQPSWDSEDSSNFKDVGSKGKVGRLSAPGSQREGRNPQLNQQ   148
```

Sequence name: /tmp/1me9ldnvfv/KbP5io8PtU:GRP_HUMAN

Sequence documentation:

Alignment of: HUMGRP5E_P5 x GRP_HUMAN ..

Alignment segment 1/1:

Quality: 1248.00

Escore: 0

Matching length: 127 Total length: 127

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLM 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

335

```
  1 MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLM 50
                .         .         .         .         .
 51 GKKSTGESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQ 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 51 GKKSTGESSSVSERGSLKQQLREYIRWEEAARNLLGLIEAKENRNHQPPQ 100
                .         .
101 PKALGNQQPSWDSEDSSNFKDVGSKGK                         127
    |||||||||||||||||||||||||||
101 PKALGNQQPSWDSEDSSNFKDVGSKGK                         127
```

Expression of GRP_HUMAN - gastrin-releasing peptide (HUMGRP5E) transcripts which are detectable by amplicon as depicted in sequence name HUMGRP5Ejunc3-7 in normal and cancerous lung tissues

Expression of GRP_HUMAN - gastrin-releasing peptide transcripts detectable by or according to HUMGRP5Ejunc3-7 amplicon (SEQ ID NO: 1648) and HUMGRP5Ejunc3-7F (SEQ ID NO: 1646) and HUMGRP5Ejunc3-7R (SEQ ID NO: 1647) primers was measured by real time PCR. In parallel the expression of four housekeeping genes PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing sample",), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 19 is a histogram showing over expression of the above-indicated GRP_HUMAN - gastrin-releasing peptide transcripts in several cancerous lung samples relative to the normal samples. As is evident from Figure 19, the expression of GRP_HUMAN - gastrin-releasing peptide transcripts detectable by the above amplicon in several cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing sample"). Notably an over-expression of at least 10 fold was found in 2 out of 15 adenocarcinoma samples, and in 7 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: HUMGRP5Ejunc3-7F forward primer; and HUMGRP5Ejunc3-7R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon

was obtained as a non-limiting illustrative example only of a suitable amplicon: HUMGRP5Ejunc3-7.

HUMGRP5Ejunc3-7F (SEQ ID NO: 1646)

ACCAGCCACCTCAACCCA

HUMGRP5Ejunc3-7R (SEQ ID NO: 1647)

CTGGAGCAGAGAGTCTTTGCCT

HUMGRP5Ejunc3-7 (SEQ ID NO: 1648)

ACCAGCCACCTCAACCCAAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAG GATAGCAGCAACTTCAAAGATGTAGGTTCAAAAGGCAAAGACTCTCTGCTCCAG

Expression of GRP_HUMAN - gastrin-releasing peptide (HUMGRP5E) transcripts which are detectable by amplicon as depicted in sequence name HUMGRP5Ejunc3-7 in different normal tissues

Expression of GRP_HUMAN - gastrin-releasing peptide transcripts detectable byor according to *HUMGRP5E junc3-7* amplicon (SEQ ID NO: 1648) and *HUMGRP5E junc3-7F* (SEQ ID NO: 1646) *and HUMGRP5E junc3-7R* (SEQ ID NO: 1647) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the breast samples (Sample Nos. 33-35,Table 3, "Tissue samples on normal panel", above), to obtain a value of relative expression of each sample relative to median of the breast samples.

HUMGRP5Ejunc3-7F (SEQ ID NO: 1646)

ACCAGCCACCTCAACCCA

HUMGRP5Ejunc3-7R (SEQ ID NO: 1647)

CTGGAGCAGAGAGTCTTTGCCT

HUMGRP5Ejunc3-7 (SEQ ID NO: 1648)

ACCAGCCACCTCAACCCAAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAG GATAGCAGCAACTTCAAAGATGTAGGTTCAAAAGGCAAAGACTCTCTGCTCCAG

The results are shown in Figure 20, demonstrating the expression of GRP_HUMAN - gastrin-releasing peptide (HUMGRP5E) transcripts which are detectable by amplicon as depicted in sequence name HUMGRP5Ejunc3-7 in different normal tissues.

DESCRIPTION FOR CLUSTER D56406

Cluster D56406 features 3 transcript(s) and 10 segment(s) of interest, the names for which are given in Tables 174 and 175, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 176.

Table 174 - Transcripts of interest

Transcript Name	Sequence ID No.
D56406_PEA_1_T3	22
D56406_PEA_1_T6	23
D56406_PEA_1_T7	24

Table 175 - Segments of interest

Segment Name	Sequence ID No.
D56406_PEA_1_node_0	340
D56406_PEA_1_node_13	341
D56406_PEA_1_node_11	342

D56406_PEA_1_node_2	343
D56406_PEA_1_node_3	344
D56406_PEA_1_node_5	345
D56406_PEA_1_node_6	346
D56406_PEA_1_node_7	347
D56406_PEA_1_node_8	348
D56406_PEA_1_node_9	349

Table 176 - Proteins of interest

Protein Name	Sequence ID No.
D56406_PEA_1_P2	1301
D56406_PEA_1_P5	1302
D56406_PEA_1_P6	1303

These sequences are variants of the known protein Neurotensin/neuromedin N precursor [Contains: Large neuromedin N (NmN- 125); Neuromedin N (NmN) (NN); Neurotensin (NT); Tail peptide] (SwissProt accession identifier NEUT_HUMAN), SEQ ID NO: 1422, referred to herein as the previously known protein.

Protein Neurotensin/neuromedin N precursor is known or believed to have the following function(s): Neurotensin may play an endocrine or paracrine role in the regulation of fat metabolism. It causes contraction of smooth muscle. The sequence for protein Neurotensin/neuromedin N precursor is given at the end of the application, as "Neurotensin/neuromedin N precursor [Contains: Large neuromedin N (NmN- 125); Neuromedin N (NmN) (NN); Neurotensin (NT); Tail peptide] amino acid sequence". Protein Neurotensin/neuromedin N precursor localization is believed to be Secreted; Packaged within secretory vesicles.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: signal transduction, which are annotation(s) related to Biological Process; neuropeptide hormone, which are annotation(s) related to Molecular Function; and extracellular; soluble fraction, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster D56406 features 3 transcript(s), which were listed in Table 174 above. These transcript(s) encode for protein(s) which are variant(s) of protein Neurotensin/neuromedin N precursor. A description of each variant protein according to the present invention is now provided.

Variant protein D56406_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) D56406_PEA_1_T3. An alignment is given to the known protein (Neurotensin/neuromedin N precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between D56406_PEA_1_P2 and NEUT_HUMAN:

1.An isolated chimeric polypeptide encoding for D56406_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAHVPSWKMT LLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEAMLTIYQLHKICHSRAF QHWE corresponding to amino acids 1 - 120 of NEUT_HUMAN, which also corresponds to amino acids 1 - 120 of D56406_PEA_1_P2, second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ARWLTPVIPALWEAETGGSRGQEMETIPANT corresponding to amino acids 121 - 151 of D56406_PEA_1_P2, and a third amino acid sequence being at least 90 % homologous to LIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 121 - 170 of NEUT_HUMAN, which also corresponds to amino acids 152 - 201 of D56406_PEA_1_P2, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for an edge portion of D56406_PEA_1_P2, comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence encoding for ARWLTPVIPALWEAETGGSRGQEMETIPANT, corresponding to D56406_PEA_1_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein D56406_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 177, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 177 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
30	M -> V	No
44	S -> P	No
84	V ->	No
84	V -> A	No

Variant protein D56406_PEA_1_P2 is encoded by the following transcript(s): D56406_PEA_1_T3, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript D56406_PEA_1_T3 is shown in bold; this coding portion starts at position 106 and ends at position 708. The transcript also has the following SNPs as listed in

Table 178 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 178 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
94	G -> T	No
95	A -> T	No
858	T -> G	Yes
103	A -> G	Yes
193	A -> G	No
235	T -> C	No
339	T -> C	No
356	T ->	No
356	T -> C	No
417	A -> T	No
757	T ->	No

Variant protein D56406_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) D56406_PEA_1_T6. An alignment is given to the known protein (Neurotensin/neuromedin N precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between D56406_PEA_1_P5 and NEUT_HUMAN:

1.An isolated chimeric polypeptide encoding for D56406_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLC

corresponding to amino acids 1 - 23 of NEUT_HUMAN, which also corresponds to amino acids 1 - 23 of D56406_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to SEEEMKALEADFLTNMHTSKISKAHVPSWKMTLLNVCSLVNNLNSPAEETGEVHEEEL VARRKLPTALDGFSLEAMLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKR KIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 26 - 170 of NEUT_HUMAN, which also corresponds to amino acids 24 - 168 of D56406_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of D56406_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise CS, having a structure as follows: a sequence starting from any of amino acid numbers 23-x to 23; and ending at any of amino acid numbers 24 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein D56406_PEA_1_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 179, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 179 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
28	M -> V	No
42	S -> P	No
82	V ->	No
82	V -> A	No

Variant protein D56406_PEA_1_P5 is encoded by the following transcript(s): D56406_PEA_1_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript D56406_PEA_1_T6 is shown in bold; this coding portion starts at position 106 and ends at position 609. The transcript also has the following SNPs as listed in Table 180 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 180 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
94	G -> T	No
95	A -> T	No
759	T -> G	Yes
806	G -> A	Yes
1014	T -> G	No
1178	T -> G	No
103	A -> G	Yes
187	A -> G	No
229	T -> C	No
333	T -> C	No
350	T ->	No

350	T -> C	No
411	A -> T	No
658	T ->	No

Variant protein D56406_PEA_1_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) D56406_PEA_1_T7. An alignment is given to the known protein (Neurotensin/neuromedin N precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between D56406_PEA_1_P6 and NEUT_HUMAN:

1.An isolated chimeric polypeptide encoding for D56406_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSK corresponding to amino acids 1 - 45 of NEUT_HUMAN, which also corresponds to amino acids 1 - 45 of D56406_PEA_1_P6, and a second amino acid sequence being at least 90 % homologous to LIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYYY corresponding to amino acids 121 - 170 of NEUT_HUMAN, which also corresponds to amino acids 46 - 95 of D56406_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of D56406_PEA_1_P6, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KL, having a structure as follows: a sequence starting from any of amino acid numbers 45-x to 45; and ending at any of amino acid numbers 46 + ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein D56406_PEA_1_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 181, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 181 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
30	M -> V	No
44	S -> P	No

Variant protein D56406_PEA_1_P6 is encoded by the following transcript(s): D56406_PEA_1_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript D56406_PEA_1_T7 is shown in bold; this coding portion starts at position 106 and ends at position 390. The transcript also has the following SNPs as listed in Table 182 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein D56406_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 182 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?

94	G -> T	No
95	A -> T	No
103	A -> G	Yes
193	A -> G	No
235	T -> C	No
439	T ->	No
540	T -> G	Yes
587	G -> A	Yes
795	T -> G	No
959	T -> G	No

As noted above, cluster D56406 features 10 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster D56406_PEA_1_node_0 according to the present invention is supported by 48 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3, D56406_PEA_1_T6 and D56406_PEA_1_T7. Table 183 below describes the starting and ending position of this segment on each transcript.

Table 183 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	1	178
D56406_PEA_1_T6	1	178
D56406_PEA_1_T7	1	178

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially

expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (with regard to lung cancer), shown in Table 184.

Table 184 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
D56406_0_5_0	lung malignant tumors	LUN

Segment cluster D56406_PEA_1_node_13 according to the present invention is supported by 43 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3, D56406_PEA_1_T6 and D56406_PEA_1_T7. Table 185 below describes the starting and ending position of this segment on each transcript.

Table 185 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	559	902
D56406_PEA_1_T6	460	1239
D56406_PEA_1_T7	241	1020

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster D56406_PEA_1_node_11 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3. Table 186 below describes the starting and ending position of this segment on each transcript.

Table 186 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	466	558

Segment cluster D56406_PEA_1_node_2 according to the present invention can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T7. Table 187 below describes the starting and ending position of this segment on each transcript.

Table 187 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	179	184
D56406_PEA_1_T7	179	184

Segment cluster D56406_PEA_1_node_3 according to the present invention is supported by 46 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3, D56406_PEA_1_T6 and D56406_PEA_1_T7. Table 188 below describes the starting and ending position of this segment on each transcript.

Table 188 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	185	240
D56406_PEA_1_T6	179	234
D56406_PEA_1_T7	185	240

Segment cluster D56406_PEA_1_node_5 according to the present invention is supported by 48 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T6. Table 189 below describes the starting and ending position of this segment on each transcript.

Table 189 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	241	355
D56406_PEA_1_T6	235	349

Segment cluster D56406_PEA_1_node_6 according to the present invention is supported by 34 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T6. Table 190 below describes the starting and ending position of this segment on each transcript.

Table 190 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	356	389
D56406_PEA_1_T6	350	383

Segment cluster D56406_PEA_1_node_7 according to the present invention is supported by 32 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T6. Table 191 below describes the starting and ending position of this segment on each transcript.

Table 191 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	390	415
D56406_PEA_1_T6	384	409

Segment cluster D56406_PEA_1_node_8 according to the present invention can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T6. Table 192 below describes the starting and ending position of this segment on each transcript.

Table 192 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	416	423
D56406_PEA_1_T6	410	417

Segment cluster D56406_PEA_1_node_9 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): D56406_PEA_1_T3 and D56406_PEA_1_T6. Table 193 below describes the starting and ending position of this segment on each transcript.

Table 193 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
D56406_PEA_1_T3	424	465
D56406_PEA_1_T6	418	459

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/jU49325aMA/8F0XuN7La5:NEUT_HUMAN

Sequence documentation:

Alignment of: D56406_PEA_1_P2 x NEUT_HUMAN    ..

Alignment segment 1/1:

                    Quality: 1591.00

Escore:        0
```

Matching length: 170 Total length: 201

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 84.58 Total Percent Identity: 84.58

Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAH 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAH 50
                .         .         .         .         .
     51 VPSWKMTLLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEA 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 VPSWKMTLLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEA 100
                .         .         .         .         .
    101 MLTIYQLHKICHSRAFQHWEARWLTPVIPALWEAETGGSRGQEMETIPAN 150
        ||||||||||||||||||
    101 MLTIYQLHKICHSRAFQHWE.............................. 120
                .         .         .         .         .
    151 TLIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYY 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    121 .LIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYY 169

    201 Y                                                 201
        |
    170 Y                                                 170
```

Sequence name: /tmp/wWui8Kd4y9/zbf3ihRwnR:NEUT_HUMAN

Sequence documentation:

Alignment of: D56406_PEA_1_P5 x NEUT_HUMAN ..

Alignment segment 1/1:

Quality: 1572.00

Escore: 0

Matching length: 168 Total length: 170

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 98.82 Total Percent Identity: 98.82

Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MMAGMKIQLVCMLLLAFSSWSLC..SEEEMKALEADFLTNMHTSKISKAH 48
          ||||||||||||||||||||||  |||||||||||||||||||||||||
        1 MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAH 50
                .         .         .         .         .
       49 VPSWKMTLLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEA 98
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       51 VPSWKMTLLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEA 100
```

```
                .         .         .         .         .
 99 MLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKRKIPYILKR 148
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 MLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKRKIPYILKR 150
                .         .
149 QLYENKPRRPYILKRDSYYY                                168
    ||||||||||||||||||||
151 QLYENKPRRPYILKRDSYYY                                170
```

Sequence name: /tmp/f5d07fF5D7/E4N5xjUIAN:NEUT_HUMAN

Sequence documentation:

Alignment of: D56406_PEA_1_P6 x NEUT_HUMAN ..

Alignment segment 1/1:

Quality: 844.00

Escore: 0

Matching length: 95 Total length: 170

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 55.88 Total Percent Identity: 55.88

Gaps: 1

```
Alignment:

                 .         .         .         .         .
         1 MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSK..... 45
           |||||||||||||||||||||||||||||||||||||||||||||||||
         1 MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAH 50
                 .         .         .         .         .
        45 ............................................... 45

        51 VPSWKMTLLNVCSLVNNLNSPAEETGEVHEEELVARRKLPTALDGFSLEA 100
                 .         .         .         .         .
        46 .....................LIQEDILDTGNDKNGKEEVIKRKIPYILKR 75
                               |||||||||||||||||||||||||||||
       101 MLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKRKIPYILKR 150
                 .         .
        76 QLYENKPRRPYILKRDSYYY                               95
           ||||||||||||||||||||
       151 QLYENKPRRPYILKRDSYYY                               170
```

DESCRIPTION FOR CLUSTER F05068

Cluster F05068 features 3 transcript(s) and 12 segment(s) of interest, the names for which are given in Tables 194and 195, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 196.

Table 194 - Transcripts of interest

Transcript Name	Sequence ID No.
F05068_PEA_1_T3	25

F05068_PEA_1_T4	26
F05068_PEA_1_T6	27

Table 195 - Segments of interest

Segment Name	Sequence ID No.
F05068_PEA_1_node_0	350
F05068_PEA_1_node_10	351
F05068_PEA_1_node_12	352
F05068_PEA_1_node_13	353
F05068_PEA_1_node_4	354
F05068_PEA_1_node_8	355
F05068_PEA_1_node_11	356
F05068_PEA_1_node_3	357
F05068_PEA_1_node_5	358
F05068_PEA_1_node_6	359
F05068_PEA_1_node_7	360
F05068_PEA_1_node_9	361

Table 196 - Proteins of interest

Protein Name	Sequence ID No.
F05068_PEA_1_P7	1304
F05068_PEA_1_P8	1305

These sequences are variants of the known protein ADM precursor [Contains: Adrenomedullin (AM); Proadrenomedullin N-20 terminal peptide (ProAM-N20) (ProAM N-terminal 20 peptide) (PAMP)] (SwissProt accession identifier ADML_HUMAN), SEQ ID NO:1423, referred to herein as the previously known protein.

Protein ADM precursor is known or believed to have the following function(s): AM and PAMP are potent hypotensive and vasodilatator agents. Numerous actions have been reported,

most related to the physiologic control of fluid and electrolyte homeostasis. In the kidney, AM is diuretic and natriuretic, and both AM and PAMP inhibit aldosterone secretion by direct adrenal actions. In pituitary gland, both peptides at physiologically relevant doses inhibit basal ACTH secretion. Both peptides appear to act in brain and pituitary gland to facilitate the loss of plasma volume, actions which complement their hypotensive effects in blood vessels. The sequence for protein ADM precursor is given at the end of the application, as "ADM precursor [Contains: Adrenomedullin (AM); Proadrenomedullin N-20 terminal peptide (ProAM-N20) (ProAM N-terminal 20 peptide) (PAMP)] amino acid sequence". Known polymorphisms for this sequence are as shown in Table 197.

Table 197 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
50	S -> R (in dbSNP:5005). /FTId=VAR_014861.

Protein ADM precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: cAMP biosynthesis; progesterone biosynthesis; signal transduction; cell-cell signaling; pregnancy; excretion; circulation; response to wounding, which are annotation(s) related to Biological Process; ligand; hormone, which are annotation(s) related to Molecular Function; and extracellular space; soluble fraction, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster F05068 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 21 refer to weighted expression of ESTs in

each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 21 and Table 198. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: uterine malignancies.

Table 198 - Normal tissue distribution

Name of Tissue	Number
bladder	164
bone	259
brain	26
colon	66
epithelial	73
general	67
head and neck	0
kidney	49
liver	0
lung	51
lymph nodes	0
breast	87
ovary	0
pancreas	30
skin	295
stomach	0
Thyroid	0
uterus	13

Table 199 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4

bladder	7.6e-01	8.0e-01	9.4e-01	0.5	9.9e-01	0.4
bone	7.5e-01	8.8e-01	1	0.1	1	0.3
brain	5.2e-01	6.1e-01	7.0e-04	2.1	1.1e-02	1.4
colon	6.2e-01	6.1e-01	9.7e-01	0.5	9.6e-01	0.6
epithelial	1.0e-01	3.0e-02	7.8e-01	0.7	5.8e-01	0.9
general	3.7e-01	2.6e-01	8.5e-01	0.8	9.0e-01	0.8
head and neck	2.1e-01	1.1e-01	1	1.0	3.2e-01	2.3
kidney	3.8e-01	3.9e-01	6.6e-02	1.8	1.2e-02	2.2
liver	1.8e-01	1.2e-01	2.3e-01	4.3	2.3e-01	2.6
lung	6.2e-01	4.3e-01	8.5e-01	0.7	3.8e-01	1.0
lymph nodes	1	3.1e-01	1	1.0	1	1.3
breast	7.8e-01	5.8e-01	9.1e-01	0.6	8.9e-01	0.7
ovary	3.8e-01	2.6e-01	3.2e-01	2.4	1.6e-01	2.5
pancreas	5.1e-01	3.3e-01	7.0e-01	0.9	1.0e-01	1.4
skin	6.0e-01	5.2e-01	9.7e-01	0.3	1	0.1
stomach	3.6e-01	3.0e-01	1	1.0	4.1e-01	1.8
Thyroid	5.0e-01	5.0e-01	6.7e-01	1.7	6.7e-01	1.7
uterus	1.1e-01	2.6e-01	2.1e-03	3.2	2.3e-02	2.2

As noted above, cluster F05068 features 3 transcript(s), which were listed in Table 194 above. These transcript(s) encode for protein(s) which are variant(s) of protein ADM precursor. A description of each variant protein according to the present invention is now provided.

Variant protein F05068_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) F05068_PEA_1_T3 and F05068_PEA_1_T6. An alignment is given to the known protein (ADM precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between F05068_PEA_1_P7 and ADML_HUMAN:

1.An isolated chimeric polypeptide encoding for F05068_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MKLVSVALMYLGSLAFLGADTARLDVASEFRKK corresponding to amino acids 1 - 33 of ADML_HUMAN, which also corresponds to amino acids 1 - 33 of F05068_PEA_1_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein F05068_PEA_1_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 200, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein F05068_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 200 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	V -> F	No
10	Y -> C	No

Variant protein F05068_PEA_1_P7 is encoded by the following transcript(s): F05068_PEA_1_T3 and F05068_PEA_1_T6, for which the sequence(s) is/are given at the end of the application.

The coding portion of transcript F05068_PEA_1_T3 is shown in bold; this coding portion starts at position 267 and ends at position 365. The transcript also has the following SNPs as listed in Table 201 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the

presence of known SNPs in variant protein F05068_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 201 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
26	C -> T	Yes
164	T ->	No
593	G -> C	Yes
860	C ->	No
860	C -> A	No
1022	G -> A	No
1023	G -> A	No
1023	G -> C	Yes
1084	G -> A	Yes
1088	C ->	No
1088	C -> A	No
1106	C ->	No
177	T ->	No
1106	C -> A	No
1149	G ->	No
1154	C ->	No
1171	T -> G	Yes
1192	G ->	No
1224	C ->	No
1266	C ->	No
1282	C -> T	No
1381	G -> A	No
1450	T ->	No
206	C -> T	Yes

1457	T -> G	No
1534	C ->	No
1535	C ->	No
1554	A -> G	Yes
1572	A -> C	No
1572	A -> G	No
1655	A -> C	Yes
1669	T -> C	Yes
1721	C -> T	No
245	G ->	No
259	C ->	No
276	G -> T	No
295	A -> G	No
317	A -> C	Yes
566	C -> G	Yes

The coding portion of transcript F05068_PEA_1_T6 is shown in bold; this coding portion starts at position 267 and ends at position 365. The transcript also has the following SNPs as listed in Table 202 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein F05068_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 202 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
26	C -> T	Yes
164	T ->	No
593	G -> C	Yes
739	C -> G	Yes
1093	C ->	No

1093	C -> A	No
1255	G -> A	No
1256	G -> A	No
1256	G -> C	Yes
1317	G -> A	Yes
1321	C ->	No
1321	C -> A	No
177	T ->	No
1339	C ->	No
1339	C -> A	No
1382	G ->	No
1387	C ->	No
1404	T -> G	Yes
1425	G ->	No
1457	C ->	No
1499	C ->	No
1515	C -> T	No
1614	G -> A	No
206	C -> T	Yes
1683	T ->	No
1690	T -> G	No
1767	C ->	No
1768	C ->	No
1787	A -> G	Yes
1805	A -> C	No
1805	A -> G	No
1888	A -> C	Yes
1902	T -> C	Yes
1954	C -> T	No
245	G ->	No

259	C ->	No
276	G -> T	No
295	A -> G	No
317	A -> C	Yes
566	C -> G	Yes

Variant protein F05068_PEA_1_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) F05068_PEA_1_T4. An alignment is given to the known protein (ADM precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between F05068_PEA_1_P8 and ADML_HUMAN:

1.An isolated chimeric polypeptide encoding for F05068_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSSSYPTGLA DVKAGPAQTLIRPQDMKGASRSPED corresponding to amino acids 1 - 82 of ADML_HUMAN, which also corresponds to amino acids 1 - 82 of F05068_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence R corresponding to amino acids 83 - 83 of F05068_PEA_1_P8, wherein said first and second amino acid sequences are contiguous and in a sequential order.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein F05068_PEA_1_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 203, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether

the SNP is known or not; the presence of known SNPs in variant protein F05068_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 203 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	V -> F	No
50	S -> R	Yes
10	Y -> C	No

Variant protein F05068_PEA_1_P8 is encoded by the following transcript(s): F05068_PEA_1_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript F05068_PEA_1_T4 is shown in bold; this coding portion starts at position 267 and ends at position 515. The transcript also has the following SNPs as listed in Table 204 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein F05068_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 204 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
26	C -> T	Yes
164	T ->	No
443	G -> C	Yes
589	C -> G	Yes
943	C ->	No
943	C -> A	No
1105	G -> A	No
1106	G -> A	No

1106	G -> C	Yes
1167	G -> A	Yes
1171	C ->	No
1171	C -> A	No
177	T ->	No
1189	C ->	No
1189	C -> A	No
1232	G ->	No
1237	C ->	No
1254	T -> G	Yes
1275	G ->	No
1307	C ->	No
1349	C ->	No
1365	C -> T	No
1464	G -> A	No
206	C -> T	Yes
1533	T ->	No
1540	T -> G	No
1617	C ->	No
1618	C ->	No
1637	A -> G	Yes
1655	A -> C	No
1655	A -> G	No
1738	A -> C	Yes
1752	T -> C	Yes
1804	C -> T	No
245	G ->	No
259	C ->	No
276	G -> T	No
295	A -> G	No

317	A -> C	Yes
416	C -> G	Yes

As noted above, cluster F05068 features 12 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster F05068_PEA_1_node_0 according to the present invention is supported by 143 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 205 below describes the starting and ending position of this segment on each transcript.

Table 205 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	1	245
F05068_PEA_1_T4	1	245
F05068_PEA_1_T6	1	245

Segment cluster F05068_PEA_1_node_10 according to the present invention is supported by 127 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 206 below describes the starting and ending position of this segment on each transcript.

Table 206 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	749	909
F05068_PEA_1_T4	832	992

F05068_PEA_1_T6	982	1142

Segment cluster F05068_PEA_1_node_12 according to the present invention is supported by 123 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 207 below describes the starting and ending position of this segment on each transcript.

Table 207 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	986	1106
F05068_PEA_1_T4	1069	1189
F05068_PEA_1_T6	1219	1339

Segment cluster F05068_PEA_1_node_13 according to the present invention is supported by 181 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 208 below describes the starting and ending position of this segment on each transcript.

Table 208 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	1107	1737
F05068_PEA_1_T4	1190	1820
F05068_PEA_1_T6	1340	1970

Segment cluster F05068_PEA_1_node_4 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): F05068_PEA_1_T3 and F05068_PEA_1_T6. Table 209 below describes the starting and ending position of this segment on each transcript.

Table 209- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	365	514
F05068_PEA_1_T6	365	514

Segment cluster F05068_PEA_1_node_8 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 210 below describes the starting and ending position of this segment on each transcript.

Table 210 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T4	515	747
F05068_PEA_1_T6	665	897

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster F05068_PEA_1_node_11 according to the present invention is supported by 112 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 211 below describes the starting and ending position of this segment on each transcript.

Table 211 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	910	985
F05068_PEA_1_T4	993	1068

F05068_PEA_1_T6	1143	1218

Segment cluster F05068_PEA_1_node_3 according to the present invention is supported by 145 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 212 below describes the starting and ending position of this segment on each transcript.

Table 212 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	246	364
F05068_PEA_1_T4	246	364
F05068_PEA_1_T6	246	364

Segment cluster F05068_PEA_1_node_5 according to the present invention is supported by 124 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 213 below describes the starting and ending position of this segment on each transcript.

Table 213 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	515	573
F05068_PEA_1_T4	365	423
F05068_PEA_1_T6	515	573

Segment cluster F05068_PEA_1_node_6 according to the present invention is supported by 110 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and

F05068_PEA_1_T6. Table 214 below describes the starting and ending position of this segment on each transcript.

Table 214 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	574	613
F05068_PEA_1_T4	424	463
F05068_PEA_1_T6	574	613

Segment cluster F05068_PEA_1_node_7 according to the present invention is supported by 109 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 215 below describes the starting and ending position of this segment on each transcript.

Table 215 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	614	664
F05068_PEA_1_T4	464	514
F05068_PEA_1_T6	614	664

Segment cluster F05068_PEA_1_node_9 according to the present invention is supported by 114 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): F05068_PEA_1_T3, F05068_PEA_1_T4 and F05068_PEA_1_T6. Table 216 below describes the starting and ending position of this segment on each transcript.

Table 216 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
F05068_PEA_1_T3	665	748

F05068_PEA_1_T4	748	831
F05068_PEA_1_T6	898	981

Variant protein alignment to the previously known protein:

Sequence name: /tmp/kEsi3RWsCN/1svdhjfiNV:ADML_HUMAN

Sequence documentation:

Alignment of: F05068_PEA_1_P7 x ADML_HUMAN ..

Alignment segment 1/1:

Quality: 304.00

Escore: 0

Matching length: 33 Total length: 33

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . .

1 MKLVSVALMYLGSLAFLGADTARLDVASEFRKK 33

373

||||||||||||||||||||||||||||||||

1 MKLVSVALMYLGSLAFLGADTARLDVASEFRKK 33

Sequence name: /tmp/tcrlWIx4kg/aghbr8Eh8n:ADML_HUMAN

Sequence documentation:

Alignment of: F05068_PEA_1_P8 x ADML_HUMAN ..

Alignment segment 1/1:

Quality: 791.00

Escore: 0

Matching length: 82 Total length: 82

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

1 MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSS 50

||||||||||||||||||||||||||||||||||||||||||||||||||

1 MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSS 50

374

```
                .          .          .
51 SYPTGLADVKAGPAQTLIRPQDMKGASRSPED                82
   ||||||||||||||||||||||||||||||||||
51 SYPTGLADVKAGPAQTLIRPQDMKGASRSPED                82
```

DESCRIPTION FOR CLUSTER H14624

Cluster H14624 features 1 transcript(s) and 15 segment(s) of interest, the names for which are given in Tables 217 and 218, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 219.

Table 217 - Transcripts of interest

Transcript Name	Sequence ID No.
H14624_T20	28

Table 218- Segments of interest

Segment Name	Sequence ID No.
H14624_node_0	362
H14624_node_16	363
H14624_node_3	364
H14624_node_10	365
H14624_node_11	366
H14624_node_12	367
H14624_node_13	368
H14624_node_14	370
H14624_node_15	371
H14624_node_4	372
H14624_node_5	373
H14624_node_6	374

H14624_node_7	375
H14624_node_8	376
H14624_node_9	377

Table 219 - Proteins of interest

Protein Name	Sequence ID No.
H14624_P15	1306

Cluster H14624 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 22 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 22 and Table 220. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: colorectal cancer, epithelial malignant tumors, a mixture of malignant tumors from different tissues, lung malignant tumors and pancreas carcinoma.

Table 220 - Normal tissue distribution

Name of Tissue	Number
adrenal	0
bladder	410
bone	71
brain	42
colon	6
epithelial	91

general	74
head and neck	0
kidney	0
lung	30
breast	949
ovary	7
pancreas	2
prostate	94
stomach	3
Thyroid	128
uterus	54

Table 221 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	4.2e-01	4.6e-01	4.6e-01	2.2	5.3e-01	1.9
bladder	5.4e-01	6.0e-01	1.2e-02	1.6	2.2e-01	1.0
bone	4.9e-01	8.5e-01	1.8e-01	1.3	7.5e-01	0.6
brain	4.7e-01	7.0e-01	6.3e-05	2.3	9.4e-03	1.4
colon	4.4e-02	9.9e-02	4.5e-03	5.4	2.0e-02	3.9
epithelial	7.7e-03	3.6e-01	1.5e-11	2.0	2.9e-02	1.1
general	5.1e-03	5.9e-01	8.3e-21	2.2	1.5e-04	1.2
head and neck	1.4e-01	2.8e-01	4.6e-01	2.2	7.5e-01	1.3
kidney	6.5e-01	7.2e-01	5.8e-01	1.7	7.0e-01	1.4
lung	6.1e-02	1.4e-01	3.3e-05	5.8	8.1e-03	2.9
breast	2.4e-01	4.1e-01	1	0.3	1	0.2
ovary	8.5e-01	7.3e-01	6.8e-01	1.2	1.6e-01	1.6
pancreas	7.5e-03	4.9e-02	1.2e-21	22.4	2.4e-16	15.1
prostate	8.3e-01	8.9e-01	7.2e-01	0.8	8.8e-01	0.6
stomach	4.6e-01	8.5e-01	1.0e-03	2.7	1.1e-01	1.4

Thyroid	7.0e-01	7.0e-01	5.9e-01	1.0	5.9e-01	1.0
uterus	4.1e-01	7.3e-01	2.3e-01	1.2	6.2e-01	0.7

As noted above, contig H14624 features 1 transcript(s), which were listed in Table 217 above. A description of each variant protein according to the present invention is now provided.

Variant protein H14624_P15 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H14624_T20. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between H14624_P15 and Q9HAP5 (SEQ ID NO:1701):

1.An isolated chimeric polypeptide encoding for H14624_P15, comprising a first amino acid sequence being at least 90 % homologous to MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLCHGIEYQNMR LPNLLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFAPVCLDDLDETIQPCHSLC VQVKDRCAPVMSAFGFPWPDMLECDRFPQDNDLCIPLASSDHLLPATEE corresponding to amino acids 1 - 167 of Q9HAP5, which also corresponds to amino acids 1 - 167 of H14624_P15, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKPSLLLPHSLLG corresponding to amino acids 168 - 180 of H14624_P15, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of H14624_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKPSLLLPHSLLG in H14624_P15.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide

prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein H14624_P15 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 222, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H14624_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 222 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
11	L ->	No
170	P -> S	Yes
28	F ->	No
29	G ->	No
38	S ->	No
45	A -> V	Yes
60	L ->	No

Variant protein H14624_P15 is encoded by the following transcript(s): H14624_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H14624_T20 is shown in bold; this coding portion starts at position 857 and ends at position 1396. The transcript also has the following SNPs as listed in Table 223 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H14624_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 223 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?

389	A -> G	No
476	C -> T	No
969	G ->	No
988	G -> T	Yes
990	C -> T	Yes
1034	C ->	No
1168	C -> T	Yes
1364	C -> T	Yes
488	T -> C	No
819	C -> G	Yes
851	C ->	No
887	C ->	No
922	G -> A	Yes
934	C -> T	Yes
938	T ->	No
943	C ->	No

As noted above, cluster H14624 features 15 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster H14624_node_0 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 224 below describes the starting and ending position of this segment on each transcript.

Table 224 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1	573

Segment cluster H14624_node_16 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 225 below describes the starting and ending position of this segment on each transcript.

Table 225 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1359	1745

Segment cluster H14624_node_3 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 226 below describes the starting and ending position of this segment on each transcript.

Table 226 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	574	822

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster H14624_node_10 according to the present invention can be found in the following transcript(s): H14624_T20. Table 227 below describes the starting and ending position of this segment on each transcript.

Table 227 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1070	1079

Segment cluster H14624_node_11 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 228 below describes the starting and ending position of this segment on each transcript.

Table 228 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1080	1114

Segment cluster H14624_node_12 according to the present invention can be found in the following transcript(s): H14624_T20. Table 229 below describes the starting and ending position of this segment on each transcript.

Table 229 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1115	1135

Segment cluster H14624_node_13 according to the present invention is supported by 124 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 230 below describes the starting and ending position of this segment on each transcript.

Table 230 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1136	1227

Segment cluster H14624_node_14 according to the present invention is supported by 114 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): H14624_T20. Table 231 below describes the starting and ending position of this segment on each transcript.

Table 231 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1228	1287

Segment cluster H14624_node_15 according to the present invention is supported by 124 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 232 below describes the starting and ending position of this segment on each transcript.

Table 232 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1288	1358

Segment cluster H14624_node_4 according to the present invention is supported by 65 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 233 below describes the starting and ending position of this segment on each transcript.

Table 233 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	823	892

Segment cluster H14624_node_5 according to the present invention can be found in the following transcript(s): H14624_T20. Table 234 below describes the starting and ending position of this segment on each transcript.

Table 234 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	893	903

Segment cluster H14624_node_6 according to the present invention can be found in the following transcript(s): H14624_T20. Table 235 below describes the starting and ending position of this segment on each transcript.

Table 235 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	904	927

Segment cluster H14624_node_7 according to the present invention can be found in the following transcript(s): H14624_T20. Table 236 below describes the starting and ending position of this segment on each transcript.

Table 236 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	928	934

Segment cluster H14624_node_8 according to the present invention is supported by 85 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 237 below describes the starting and ending position of this segment on each transcript.

Table 237 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	935	1014

Segment cluster H14624_node_9 according to the present invention is supported by 87 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H14624_T20. Table 238 below describes the starting and ending position of this segment on each transcript.

Table 238 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H14624_T20	1015	1069

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/Upb1SbFkrj/N4PrGQAB2V:Q9HAP5

Sequence documentation:

Alignment of: H14624_P15 x Q9HAP5    ..

Alignment segment 1/1:

                        Quality: 1702.00
Escore:         0
                Matching length:      167                  Total
length:       167
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
```

```
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                          Gaps:       0

Alignment:

                    .         .         .         .         .
          1 MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLC 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
          1 MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLC 50
                    .         .         .         .         .
         51 HGIEYQNMRLPNLLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFA 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||
         51 HGIEYQNMRLPNLLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFA 100
                    .         .         .         .         .
        101 PVCLDDLDETIQPCHSLCVQVKDRCAPVMSAFGFPWPDMLECDRFPQDND 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        101 PVCLDDLDETIQPCHSLCVQVKDRCAPVMSAFGFPWPDMLECDRFPQDND 150
                    .
        151 LCIPLASSDHLLPATEE                                  167
            |||||||||||||||||
        151 LCIPLASSDHLLPATEE                                  167
```

DESCRIPTION FOR CLUSTER H38804

Cluster H38804 features 2 transcript(s) and 20 segment(s) of interest, the names for which are given in Tables 239 and 240, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 241.

Table 239 - Transcripts of interest

Transcript Name	Sequence ID No.

H38804_PEA_1_T24	29
H38804_PEA_1_T8	30

Table 240 - Segments of interest

Segment Name	Sequence ID No.
H38804_PEA_1_node_0	378
H38804_PEA_1_node_1	379
H38804_PEA_1_node_16	380
H38804_PEA_1_node_19	381
H38804_PEA_1_node_24	382
H38804_PEA_1_node_25	383
H38804_PEA_1_node_28	384
H38804_PEA_1_node_29	385
H38804_PEA_1_node_30	386
H38804_PEA_1_node_10	387
H38804_PEA_1_node_12	388
H38804_PEA_1_node_13	389
H38804_PEA_1_node_14	390
H38804_PEA_1_node_2	391
H38804_PEA_1_node_20	392
H38804_PEA_1_node_23	393
H38804_PEA_1_node_26	394
H38804_PEA_1_node_3	395
H38804_PEA_1_node_4	396
H38804_PEA_1_node_5	397

Table 241 - Proteins of interest

Protein Name	Sequence ID No.
H38804_PEA_1_P5	1307

H38804_PEA_1_P17	1308

These sequences are variants of the known protein Mitotic checkpoint protein BUB3 (SwissProt accession identifier BUB3_HUMAN), SEQ ID NO:1424, referred to herein as the previously known protein.

Protein Mitotic checkpoint protein BUB3 is known or believed to have the following function(s): Required for kinetochore localization of BUB1. The sequence for protein Mitotic checkpoint protein BUB3 is given at the end of the application, as "Mitotic checkpoint protein BUB3 amino acid sequence". Known polymorphisms for this sequence are as shown in Table 242

Table 242 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
326 - 327	Missing

Protein Mitotic checkpoint protein BUB3 localization is believed to be Nuclear.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: mitosis; mitotic checkpoint; mitotic spindle checkpoint; cell proliferation, which are annotation(s) related to Biological Process; and nucleus, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster H38804 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 23 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 23 and Table 243. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: transitional cell carcinoma, brain malignant tumors, a mixture of malignant tumors from different tissues and gastric carcinoma.

Table 243 - Normal tissue distribution

Name of Tissue	Number
adrenal	124
bladder	0
bone	64
brain	40
colon	75
epithelial	86
general	79
head and neck	334
kidney	69
liver	14
lung	125
lymph nodes	218
breast	263
bone marrow	62
muscle	27
ovary	109
pancreas	43
prostate	32
skin	53
stomach	0
T cells	557
Thyroid	257

uterus	113

Table 244 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	6.3e-01	5.4e-01	1.8e-01	1.4	5.0e-02	1.9
bladder	7.0e-02	2.6e-02	3.2e-02	4.9	9.9e-03	6.2
bone	3.7e-01	2.3e-01	7.9e-01	0.9	3.2e-01	1.6
brain	3.1e-02	4.2e-03	5.3e-01	1.2	1.1e-02	2.1
colon	2.4e-01	1.1e-01	2.0e-01	1.7	1.6e-01	1.8
epithelial	1.1e-01	2.2e-02	1.5e-01	1.2	8.6e-03	1.3
general	2.3e-02	2.3e-04	9.0e-02	1.2	4.7e-05	1.4
head and neck	4.4e-01	4.7e-01	9.2e-01	0.6	8.9e-01	0.5
kidney	8.2e-01	8.4e-01	9.0e-01	0.8	3.5e-01	1.0
liver	8.3e-01	1.5e-01	1	0.8	5.3e-02	2.8
lung	6.9e-01	8.1e-01	5.1e-01	1.1	6.0e-01	0.8
lymph nodes	5.1e-01	6.9e-01	5.0e-01	0.9	9.5e-01	0.5
breast	4.9e-01	4.2e-01	9.7e-01	0.5	9.5e-01	0.5
bone marrow	6.7e-01	5.4e-01	1	1.5	3.3e-02	2.6
muscle	8.5e-01	6.1e-01	1	0.4	6.3e-01	1.0
ovary	3.4e-01	3.3e-01	2.5e-01	1.5	4.7e-01	1.1
pancreas	4.3e-01	4.9e-01	6.3e-01	1.0	6.9e-01	0.9
prostate	7.4e-01	6.5e-01	1.5e-01	1.9	1.0e-01	2.0
skin	6.0e-01	1.7e-01	5.4e-01	1.4	2.7e-02	1.2
stomach	4.5e-02	9.9e-03	2.5e-01	3.1	4.3e-02	4.3
T cells	5.0e-01	6.7e-01	1	0.3	9.8e-01	0.5
Thyroid	5.7e-01	5.7e-01	1	0.4	1	0.4
uterus	5.7e-01	6.7e-01	9.2e-01	0.6	8.7e-01	0.5

As noted above, cluster H38804 features 2 transcript(s), which were listed in Table 239 above. These transcript(s) encode for protein(s) which are variant(s) of protein Mitotic

checkpoint protein BUB3. A description of each variant protein according to the present invention is now provided.

Variant protein H38804_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H38804_PEA_1_T8. An alignment is given to the known protein (Mitotic checkpoint protein BUB3) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between H38804_PEA_1_P5 and BUB3_HUMAN:

1.An isolated chimeric polypeptide encoding for H38804_PEA_1_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ corresponding to amino acids 1 - 57 of H38804_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGA VLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIRCVEYCPEVNVMVTG SWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQ QRRESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENN IEQIYPVNAISFHNIHNTFATGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTL AIASSYMYEMDDTEHPEDGIFIRQVTDAETKPK corresponding to amino acids 1 - 324 of BUB3_HUMAN, which also corresponds to amino acids 58 - 381 of H38804_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of H38804_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ of H38804_PEA_1_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Signal peptide,NN:NO) predicts that this protein has a signal peptide..

Variant protein H38804_PEA_1_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 245, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H38804_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 245 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
126	H -> Y	No
129	S -> R	Yes
256	I ->	No
256	I -> N	No
258	G ->	No
266	D ->	No
266	D -> E	No
266	D -> N	Yes
296	A -> G	No
296	A -> V	No
306	F -> C	No
314	F ->	No

215	R -> K	No
361	T -> A	No
381	K ->	No
217	L ->	No
220	D ->	No
220	D -> E	No
245	F ->	No
245	F -> V	No
248	K ->	No
248	K -> Q	No

Variant protein H38804_PEA_1_P5 is encoded by the following transcript(s): H38804_PEA_1_T8, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H38804_PEA_1_T8 is shown in bold; this coding portion starts at position 475 and ends at position 1617. The transcript also has the following SNPs as listed in Table 246 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H38804_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 246 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
161	C ->	No
167	C ->	No
1118	G -> A	No
1123	T ->	No
1134	C ->	No
1134	C -> A	No
1207	T ->	No
1207	T -> G	No

1216	A ->	No
1216	A -> C	No
1241	T ->	No
1241	T -> A	No
167	C -> A	No
1248	C ->	No
1248	C -> G	No
1270	G -> A	Yes
1272	C ->	No
1272	C -> A	No
1361	C -> G	No
1361	C -> T	No
1391	T -> G	No
1414	T ->	No
1419	A -> G	No
192	T ->	No
1555	A -> G	No
1615	A ->	No
1642	G -> A	Yes
1846	T -> C	Yes
2090	A -> G	No
2356	C -> G	No
2712	G ->	No
2909	T -> C	No
2909	T -> G	No
3020	T -> G	No
208	C -> T	Yes
3251	T ->	No
3306	T ->	No
3307	T -> G	No

3354	T ->	No
3521	-> G	No
3601	C ->	No
3601	C -> G	No
3633	T ->	No
3633	T -> G	No
3638	A ->	No
849	G -> T	No
3638	A -> C	No
3674	C -> T	Yes
3812	T -> G	No
3862	G -> A	Yes
3864	T -> A	No
3865	T -> A	No
3990	T -> G	No
4096	T -> G	No
4152	G -> A	Yes
850	C -> T	No
855	C -> T	Yes
861	T -> G	Yes
1098	T -> C	No

Variant protein H38804_PEA_1_P17 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H38804_PEA_1_T24. An alignment is given to the known protein (Mitotic checkpoint protein BUB3) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between H38804_PEA_1_P17 and BUB3_HUMAN:

1.An isolated chimeric polypeptide encoding for H38804_PEA_1_P17, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ corresponding to amino acids 1 - 57 of H38804_PEA_1_P17, and a second amino acid sequence being at least 90 % homologous to MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGA VLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIRCVEYCPEVNVMVTG SWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQ QRRESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENN IEQIYPVNAISFHNIHNTFATGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTL AIASSYMYEMDDTEHPEDGIFIRQVTDAETKPKSPCT corresponding to amino acids 1 - 328 of BUB3_HUMAN, which also corresponds to amino acids 58 - 385 of H38804_PEA_1_P17, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of H38804_PEA_1_P17, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQ of H38804_PEA_1_P17.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Signal peptide,NN:NO) predicts that this protein has a signal peptide..

Variant protein H38804_PEA_1_P17 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 247, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H38804_PEA_1_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 247 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
126	H -> Y	No
129	S -> R	Yes
256	I ->	No
256	I -> N	No
258	G ->	No
266	D ->	No
266	D -> E	No
266	D -> N	Yes
296	A -> G	No
296	A -> V	No
306	F -> C	No
314	F ->	No
215	R -> K	No
361	T -> A	No
381	K ->	No
217	L ->	No
220	D ->	No
220	D -> E	No
245	F ->	No
245	F -> V	No
248	K ->	No

248	K -> Q	No

Variant protein H38804_PEA_1_P17 is encoded by the following transcript(s): H38804_PEA_1_T24, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H38804_PEA_1_T24 is shown in bold; this coding portion starts at position 475 and ends at position 1629. The transcript also has the following SNPs as listed in Table 248 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H38804_PEA_1_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 248 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
161	C ->	No
167	C ->	No
1118	G -> A	No
1123	T ->	No
1134	C ->	No
1134	C -> A	No
1207	T ->	No
1207	T -> G	No
1216	A ->	No
1216	A -> C	No
1241	T ->	No
1241	T -> A	No
167	C -> A	No
1248	C ->	No
1248	C -> G	No
1270	G -> A	Yes
1272	C ->	No

1272	C -> A	No
1361	C -> G	No
1361	C -> T	No
1391	T -> G	No
1414	T ->	No
1419	A -> G	No
192	T ->	No
1555	A -> G	No
1615	A ->	No
1721	G ->	No
1918	T -> C	No
1918	T -> G	No
2029	T -> G	No
2260	T ->	No
2315	T ->	No
2316	T -> G	No
2363	T ->	No
208	C -> T	Yes
2530	-> G	No
2610	C ->	No
2610	C -> G	No
2642	T ->	No
2642	T -> G	No
2647	A ->	No
2647	A -> C	No
2683	C -> T	Yes
2821	T -> G	No
2871	G -> A	Yes
849	G -> T	No
2873	T -> A	No

2874	T -> A	No
2999	T -> G	No
3105	T -> G	No
3161	G -> A	Yes
850	C -> T	No
855	C -> T	Yes
861	T -> G	Yes
1098	T -> C	No

As noted above, cluster H38804 features 20 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster H38804_PEA_1_node_0 according to the present invention is supported by 125 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 249 below describes the starting and ending position of this segment on each transcript.

Table 249 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1	213
H38804_PEA_1_T8	1	213

Segment cluster H38804_PEA_1_node_1 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 250 below describes the starting and ending position of this segment on each transcript.

Table 250 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	214	645
H38804_PEA_1_T8	214	645

Segment cluster H38804_PEA_1_node_16 according to the present invention is supported by 214 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 251 below describes the starting and ending position of this segment on each transcript.

Table 251 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1063	1221
H38804_PEA_1_T8	1063	1221

Segment cluster H38804_PEA_1_node_19 according to the present invention is supported by 198 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 252 below describes the starting and ending position of this segment on each transcript.

Table 252 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1222	1360
H38804_PEA_1_T8	1222	1360

Segment cluster H38804_PEA_1_node_24 according to the present invention is supported by 180 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 253 below describes the starting and ending position of this segment on each transcript.

Table 253 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1421	1616
H38804_PEA_1_T8	1421	1616

Segment cluster H38804_PEA_1_node_25 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T8. Table 254 below describes the starting and ending position of this segment on each transcript.

Table 254 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T8	1617	1969

Segment cluster H38804_PEA_1_node_28 according to the present invention is supported by 38 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T8. Table 255 below describes the starting and ending position of this segment on each transcript.

Table 255 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T8	2018	2607

Segment cluster H38804_PEA_1_node_29 according to the present invention is supported by 259 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 256 below describes the starting and ending position of this segment on each transcript.

Table 256 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1617	2844
H38804_PEA_1_T8	2608	3835

Segment cluster H38804_PEA_1_node_30 according to the present invention is supported by 169 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 257 below describes the starting and ending position of this segment on each transcript.

Table 257 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	2845	3170
H38804_PEA_1_T8	3836	4161

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster H38804_PEA_1_node_10 according to the present invention is supported by 179 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 258 below describes the starting and ending position of this segment on each transcript.

Table 258 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	841	910
H38804_PEA_1_T8	841	910

Segment cluster H38804_PEA_1_node_12 according to the present invention is supported by 181 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 259 below describes the starting and ending position of this segment on each transcript.

Table 259 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	911	949
H38804_PEA_1_T8	911	949

Segment cluster H38804_PEA_1_node_13 according to the present invention is supported by 187 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 260 below describes the starting and ending position of this segment on each transcript.

Table 260 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	950	1028
H38804_PEA_1_T8	950	1028

Segment cluster H38804_PEA_1_node_14 according to the present invention is supported by 179 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 261 below describes the starting and ending position of this segment on each transcript.

Table 261 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1029	1062
H38804_PEA_1_T8	1029	1062

Segment cluster H38804_PEA_1_node_2 according to the present invention is supported by 156 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 262 below describes the starting and ending position of this segment on each transcript.

Table 262 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	646	678
H38804_PEA_1_T8	646	678

Segment cluster H38804_PEA_1_node_20 according to the present invention is supported by 162 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 263 below describes the starting and ending position of this segment on each transcript.

Table 263 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1361	1399
H38804_PEA_1_T8	1361	1399

Segment cluster H38804_PEA_1_node_23 according to the present invention can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 264 below describes the starting and ending position of this segment on each transcript.

Table 264 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	1400	1420
H38804_PEA_1_T8	1400	1420

Segment cluster H38804_PEA_1_node_26 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T8. Table 265 below describes the starting and ending position of this segment on each transcript.

Table 265 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T8	1970	2017

Segment cluster H38804_PEA_1_node_3 according to the present invention is supported by 162 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 266 below describes the starting and ending position of this segment on each transcript.

Table 266 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	679	716
H38804_PEA_1_T8	679	716

Segment cluster H38804_PEA_1_node_4 according to the present invention is supported by 172 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 267 below describes the starting and ending position of this segment on each transcript.

Table 267 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	717	827
H38804_PEA_1_T8	717	827

Segment cluster H38804_PEA_1_node_5 according to the present invention can be found in the following transcript(s): H38804_PEA_1_T24 and H38804_PEA_1_T8. Table 268 below describes the starting and ending position of this segment on each transcript.

Table 268 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H38804_PEA_1_T24	828	840
H38804_PEA_1_T8	828	840

Variant protein alignment to the previously known protein:

Sequence name: /tmp/RR4oV8zYLg/QlORqeqpIp:BUB3_HUMAN

Sequence documentation:

Alignment of: H38804_PEA_1_P5 x BUB3_HUMAN ..

Alignment segment 1/1:

Quality: 3244.00

Escore: 0

Matching length: 324 Total length: 324

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
 58 MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMR 107
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMR 50
                .         .         .         .         .
108 LKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIR 157
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIR 100
                .         .         .         .         .
158 CVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRL 207
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 CVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRL 150
                .         .         .         .         .
208 IVGTAGRRVLVWDLRNMGYVQQRRESSLKYQTRCIRAFPNKQGYVLSSIE 257
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 IVGTAGRRVLVWDLRNMGYVQQRRESSLKYQTRCIRAFPNKQGYVLSSIE 200
                .         .         .         .         .
258 GRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFA 307
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 GRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFA 250
                .         .         .         .         .
308 TGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYE 357
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 TGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYE 300
                .         .
358 MDDTEHPEDGIFIRQVTDAETKPK                           381
```

```
                                    408
                    ||||||||||||||||||||||||
          301 MDDTEHPEDGIFIRQVTDAETKPK                                  324
```

Sequence name: /tmp/Db0dQEpSuo/Lr8HPXaeBg:BUB3_HUMAN

Sequence documentation:

Alignment of: H38804_PEA_1_P17 x BUB3_HUMAN ..

Alignment segment 1/1:

Quality: 3288.00

Escore: 0

Matching length: 328 Total length: 328

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                    .         .         .         .         .
           58 MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMR 107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
            1 MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMR 50
                    .         .         .         .         .
          108 LKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIR 157
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDAPIR 100
             .         .         .         .         .
158 CVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRL 207
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 CVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRL 150
             .         .         .         .         .
208 IVGTAGRRVLVWDLRNMGYVQQRRESSLKYQTRCIRAFPNKQGYVLSSIE 257
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 IVGTAGRRVLVWDLRNMGYVQQRRESSLKYQTRCIRAFPNKQGYVLSSIE 200
             .         .         .         .         .
258 GRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFA 307
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 GRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFA 250
             .         .         .         .         .
308 TGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYE 357
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 TGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYE 300
             .         .
358 MDDTEHPEDGIFIRQVTDAETKPKSPCT                       385
    ||||||||||||||||||||||||||||
301 MDDTEHPEDGIFIRQVTDAETKPKSPCT                       328
```

DESCRIPTION FOR CLUSTER HSENA78

Cluster HSENA78 features 1 transcript(s) and 7 segment(s) of interest, the names for which are given in Tables 269 and 270, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 271.

Table 269 - Transcripts of interest

Transcript Name	Sequence ID No.
HSENA78_T5	31

Table 270 - Segments of interest

Segment Name	Sequence ID No.
HSENA78_node_0	398
HSENA78_node_2	399
HSENA78_node_6	400
HSENA78_node_9	401
HSENA78_node_3	402
HSENA78_node_4	403
HSENA78_node_8	404

Table 271 - Proteins of interest

Protein Name	Sequence ID No.
HSENA78_P2	1309

These sequences are variants of the known protein Small inducible cytokine B5 precursor (SwissProt accession identifier SZ05_HUMAN; known also according to the synonyms CXCL5; Epithelial-derived neutrophil activating protein 78; Neutrophil-activating peptide ENA- 78), SEQ ID NO: 1425, referred to herein as the previously known protein.

Protein Small inducible cytokine B5 precursor is known or believed to have the following function(s): Involved in neutrophil activation. The sequence for protein Small inducible cytokine B5 precursor is given at the end of the application, as "Small inducible cytokine B5 precursor amino acid sequence". Protein Small inducible cytokine B5 precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: chemotaxis; signal transduction; cell-cell signaling; positive control of cell proliferation, which are annotation(s) related to Biological Process; and chemokine, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HSENA78 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 24 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 24 and Table 272. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors and lung malignant tumors.

Table 272 - Normal tissue distribution

Name of Tissue	Number
colon	0
epithelial	2
general	38
kidney	0
lung	3
breast	8
skin	0
stomach	36
uterus	4

Table 273 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
colon	2.6e-01	3.3e-01	1.7e-01	2.7	2.7e-01	2.2
epithelial	2.5e-01	9.0e-02	3.2e-03	4.1	8.5e-07	5.5
general	8.4e-01	7.2e-01	1	0.3	1	0.4
kidney	1	7.2e-01	1	1.0	1.7e-01	1.9

lung	8.5e-01	4.8e-01	4.1e-01	1.9	4.0e-05	3.8
breast	9.5e-01	8.7e-01	1	0.8	6.8e-01	1.2
skin	2.9e-01	4.7e-01	1.4e-01	7.0	6.4e-01	1.6
stomach	5.0e-01	4.3e-01	7.5e-01	1.0	4.3e-01	1.3
uterus	7.1e-01	8.5e-01	6.6e-01	1.3	8.0e-01	1.0

As noted above, cluster HSENA78 features 1 transcript(s), which were listed in Table 269 above. These transcript(s) encode for protein(s) which are variant(s) of protein Small inducible cytokine B5 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HSENA78_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSENA78_T5. An alignment is given to the known protein (Small inducible cytokine B5 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSENA78_P2 and SZ05_HUMAN:

1.An isolated chimeric polypeptide encoding for HSENA78_P2, comprising a first amino acid sequence being at least 90 % homologous to MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCVCLQTTQGVHP KMISNLQVFAIGPQCSKVEVV corresponding to amino acids 1 - 81 of SZ05_HUMAN, which also corresponds to amino acids 1 - 81 of HSENA78_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein HSENA78_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 274, (given according to their position(s) on the amino acid

sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSENA78_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 274 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
80	V ->	No
81	V ->	No

Variant protein HSENA78_P2 is encoded by the following transcript(s): HSENA78_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSENA78_T5 is shown in bold; this coding portion starts at position 149 and ends at position 391. The transcript also has the following SNPs as listed in Table 275 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSENA78_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 275 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
92	C -> T	Yes
144	C -> T	No
1151	A -> T	Yes
1389	T -> C	No
1867	C -> G	Yes
145	C -> T	No
181	C -> T	Yes
316	G -> A	Yes
388	G ->	No

390	T ->	No
605	T ->	No
972	C -> T	Yes
1105	A -> G	Yes

As noted above, cluster HSENA78 features 7 segment(s), which were listed in Table 270 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HSENA78_node_0 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSENA78_T5. Table 276 below describes the starting and ending position of this segment on each transcript.

Table 276 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	1	257

Segment cluster HSENA78_node_2 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSENA78_T5. Table 277 below describes the starting and ending position of this segment on each transcript.

Table 277 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	258	390

Segment cluster HSENA78_node_6 according to the present invention is supported by 68 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): HSENA78_T5. Table 278 below describes the starting and ending position of this segment on each transcript.

Table 278 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	585	2370

Segment cluster HSENA78_node_9 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSENA78_T5. Table 279 below describes the starting and ending position of this segment on each transcript.

Table 279 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	2394	2546

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HSENA78_node_3 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSENA78_T5. Table 280 below describes the starting and ending position of this segment on each transcript.

Table 280 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	391	500

Segment cluster HSENA78_node_4 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): HSENA78_T5. Table 281 below describes the starting and ending position of this segment on each transcript.

Table 281 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	501	584

Segment cluster HSENA78_node_8 according to the present invention can be found in the following transcript(s): HSENA78_T5. Table 282 below describes the starting and ending position of this segment on each transcript.

Table 282 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSENA78_T5	2371	2393

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/5kiQY6MxWx/pLnTrxsCqk:SZ05_HUMAN

Sequence documentation:

Alignment of: HSENA78_P2 x SZ05_HUMAN    ..

Alignment segment 1/1:

                    Quality:  767.00
Escore:        0
```

417

Matching length: 81 Total length: 81

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCV 50
                  .         .         .
       51 CLQTTQGVHPKMISNLQVFAIGPQCSKVEVV                    81
          ||||||||||||||||||||||||||||||||
       51 CLQTTQGVHPKMISNLQVFAIGPQCSKVEVV                    81
```

DESCRIPTION FOR CLUSTER HUMODCA

Cluster HUMODCA features 1 transcript(s) and 17 segment(s) of interest, the names for which are given in Tables 283 and 284, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 285.

Table 283 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMODCA_T17	32

Table 284 - Segments of interest

Segment Name	Sequence ID No.

HUMODCA_node_1	405
HUMODCA_node_25	406
HUMODCA_node_32	407
HUMODCA_node_36	408
HUMODCA_node_39	409
HUMODCA_node_41	410
HUMODCA_node_0	411
HUMODCA_node_10	412
HUMODCA_node_12	413
HUMODCA_node_13	414
HUMODCA_node_2	415
HUMODCA_node_27	416
HUMODCA_node_3	417
HUMODCA_node_30	418
HUMODCA_node_34	419
HUMODCA_node_38	420
HUMODCA_node_40	421

Table 285 - Proteins of interest

Protein Name	Sequence ID No.
HUMODCA_P9	1310

These sequences are variants of the known protein Ornithine decarboxylase (SwissProt accession identifier DCOR_HUMAN; known also according to the synonyms EC 4.1.1.17; ODC), SEQ ID NO: 1426, referred to herein as the previously known protein.

Protein Ornithine decarboxylase is known or believed to have the following function(s): Polyamine biosynthesis; first (rate-limiting) step. The sequence for protein Ornithine decarboxylase is given at the end of the application, as "Ornithine decarboxylase amino acid sequence". Known polymorphisms for this sequence are as shown in Table 286.

Table 286 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
415	Q -> E

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: polyamine biosynthesis, which are annotation(s) related to Biological Process; and ornithine decarboxylase; lyase, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HUMODCA can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 25 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 25 and Table 287. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: brain malignant tumors, colorectal cancer, epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Table 287 - Normal tissue distribution

Name of Tissue	Number
adrenal	120
bladder	82
bone	161
brain	53
colon	0
epithelial	107

general	94
head and neck	10
kidney	114
liver	107
lung	120
lymph nodes	165
breast	61
bone marrow	156
muscle	55
ovary	36
pancreas	102
prostate	140
skin	188
stomach	109
T cells	278
Thyroid	128
uterus	118

Table 288 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	8.3e-01	7.8e-01	1	0.2	8.5e-01	0.7
bladder	5.4e-01	5.1e-01	6.2e-01	1.1	5.0e-01	1.1
bone	8.3e-01	3.2e-01	1	0.2	8.4e-01	0.7
brain	2.6e-01	3.8e-02	6.5e-04	2.8	8.7e-10	3.6
colon	2.2e-02	5.8e-03	1.5e-03	6.9	6.7e-05	9.9
epithelial	6.4e-02	2.7e-03	1.4e-03	1.5	1.6e-12	2.1
general	1.3e-03	5.4e-08	1.9e-08	1.7	1.4e-39	2.6
head and neck	1.7e-01	1.7e-01	1	1.2	7.5e-01	1.3
kidney	7.7e-01	7.6e-01	7.1e-01	0.8	6.6e-01	0.9

liver	7.3e-01	5.7e-01	1	0.3	2.4e-01	1.2
lung	7.8e-01	5.8e-01	7.6e-01	0.6	7.3e-04	1.7
lymph nodes	3.9e-01	2.5e-01	1.8e-01	1.1	1.4e-04	2.1
breast	7.8e-01	4.7e-01	7.7e-01	0.8	6.4e-01	1.0
bone marrow	3.4e-01	2.6e-01	2.8e-01	2.1	1.6e-01	1.2
muscle	8.5e-01	6.1e-01	1	0.2	7.1e-05	1.0
ovary	1.7e-01	9.3e-02	3.8e-01	1.7	2.2e-02	2.6
pancreas	2.2e-01	3.2e-01	5.7e-02	1.6	6.6e-03	1.5
prostate	5.0e-01	4.9e-01	3.8e-02	1.9	4.5e-02	1.7
skin	6.2e-01	5.8e-01	5.4e-02	0.9	1.5e-02	0.5
stomach	4.2e-01	2.6e-01	3.7e-01	0.7	7.3e-03	2.3
T cells	1	1	5.5e-01	1.5	8.1e-01	0.9
Thyroid	8.3e-02	8.3e-02	5.9e-01	1.3	5.9e-01	1.3
uterus	4.2e-01	2.4e-01	1.6e-01	1.2	4.9e-02	1.7

As noted above, cluster HUMODCA features 1 transcript(s), which were listed in Table 283 above. These transcript(s) encode for protein(s) which are variant(s) of protein Ornithine decarboxylase. A description of each variant protein according to the present invention is now provided.

Variant protein HUMODCA_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMODCA_T17. An alignment is given to the known protein (Ornithine decarboxylase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMODCA_P9 and DCOR_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to

LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV
QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG
VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN
CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN
MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA
WESGMKRHRAACASASINV corresponding to amino acids 151 - 461 of DCOR_HUMAN, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

Comparison report between HUMODCA_P9 and AAA59968 (SEQ ID NO:1702):

1.An isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA WESGMKRHRAACASASINV corresponding to amino acids 40 - 350 of AAA59968, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

Comparison report between HUMODCA_P9 and AAH14562 (SEQ ID NO:1703):

1.An isolated chimeric polypeptide encoding for HUMODCA_P9, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL corresponding to amino acids 1 - 29 of HUMODCA_P9, and a second amino acid sequence being at least 90 % homologous to LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFV QAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSG VRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFN CILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCA WESGMKRHRAACASASINV corresponding to amino acids 86 - 396 of AAH14562, which also corresponds to amino acids 30 - 340 of HUMODCA_P9, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of HUMODCA_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MKSLTATSSMKVLLPRTFWTRKLMKFLLL of HUMODCA_P9.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMODCA_P9 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 289, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMODCA_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 289 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
150	I -> S	No
150	I -> V	No
262	F -> L	No
263	E ->	No
263	E -> G	No
30	L ->	No
301	N ->	No
301	N -> K	No
309	E -> K	No
312	D -> N	No
323	E -> K	No
329	H -> P	No
174	I ->	No
34	I ->	No
59	L ->	No
70	V ->	No
86	T ->	No
86	T -> N	No
90	A ->	No
94	A ->	No
97	V ->	No
97	V -> G	No
198	N -> D	No
200	G ->	No
3	S ->	No
207	C -> G	No
207	C -> R	No
223	P ->	No

262	F ->	No

Variant protein HUMODCA_P9 is encoded by the following transcript(s): HUMODCA_T17, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMODCA_T17 is shown in bold; this coding portion starts at position 528 and ends at position 1547. The transcript also has the following SNPs as listed in Table 290 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMODCA_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 290 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
28	C -> G	Yes
210	C ->	No
536	T ->	No
615	T ->	No
628	T ->	No
703	T ->	No
736	T ->	No
784	C ->	No
784	C -> A	No
797	A ->	No
797	A -> T	No
808	C ->	No
217	C ->	No
817	T ->	No
817	T -> G	No
869	C -> T	Yes
975	A -> G	No

976	T -> G	No
1048	T ->	No
1119	A -> G	No
1127	C ->	No
1127	C -> G	No
1146	T -> C	No
366	G -> C	No
1146	T -> G	No
1194	C ->	No
1283	T -> C	Yes
1311	T ->	No
1311	T -> C	No
1315	A ->	No
1315	A -> G	No
1430	C ->	No
1430	C -> A	No
1433	C -> G	No
366	G -> T	No
1433	C -> T	Yes
1452	G -> A	No
1461	G -> A	No
1494	G -> A	No
1513	A -> C	No
1632	T ->	No
1673	C ->	No
1739	T ->	No
1739	T -> G	No
1742	T -> C	No
447	G -> A	Yes
1786	C ->	No

1786	C -> G	No
1832	T -> C	Yes
1877	C -> T	No
464	T -> G	Yes
473	A -> G	Yes
506	G -> A	Yes
521	T ->	No

As noted above, cluster HUMODCA features 17 segment(s), which were listed in Table 284 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMODCA_node_1 according to the present invention is supported by 76 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 291 below describes the starting and ending position of this segment on each transcript.

Table 291 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	118	256

Segment cluster HUMODCA_node_25 according to the present invention is supported by 190 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 292 below describes the starting and ending position of this segment on each transcript.

Table 292 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	614	748

Segment cluster HUMODCA_node_32 according to the present invention is supported by 249 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 293 below describes the starting and ending position of this segment on each transcript.

Table 293 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	915	1077

Segment cluster HUMODCA_node_36 according to the present invention is supported by 348 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 294 below describes the starting and ending position of this segment on each transcript.

Table 294 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1191	1405

Segment cluster HUMODCA_node_39 according to the present invention is supported by 297 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 295 below describes the starting and ending position of this segment on each transcript.

Table 295 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1461	1633

Segment cluster HUMODCA_node_41 according to the present invention is supported by 230 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 296 below describes the starting and ending position of this segment on each transcript.

Table 296 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1728	1893

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMODCA_node_0 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 297 below describes the starting and ending position of this segment on each transcript.

Table 297 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1	117

Segment cluster HUMODCA_node_10 according to the present invention is supported by 107 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 298 below describes the starting and ending position of this segment on each transcript.

Table 298 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	385	494

Segment cluster HUMODCA_node_12 according to the present invention is supported by 132 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 299 below describes the starting and ending position of this segment on each transcript.

Table 299 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	495	586

Segment cluster HUMODCA_node_13 according to the present invention is supported by 126 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 300 below describes the starting and ending position of this segment on each transcript.

Table 300 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	587	613

Segment cluster HUMODCA_node_2 according to the present invention is supported by 81 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 301 below describes the starting and ending position of this segment on each transcript.

Table 301 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	257	328

Segment cluster HUMODCA_node_27 according to the present invention is supported by 185 libraries. The number of libraries was determined as previously described. This segment can

be found in the following transcript(s): HUMODCA_T17. Table 302 below describes the starting and ending position of this segment on each transcript.

Table 302 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	749	830

Segment cluster HUMODCA_node_3 according to the present invention is supported by 85 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 303 below describes the starting and ending position of this segment on each transcript.

Table 303 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	329	384

Segment cluster HUMODCA_node_30 according to the present invention is supported by 196 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 304 below describes the starting and ending position of this segment on each transcript.

Table 304 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	831	914

Segment cluster HUMODCA_node_34 according to the present invention is supported by 259 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 305 below describes the starting and ending position of this segment on each transcript.

Table 305 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1078	1190

Segment cluster HUMODCA_node_38 according to the present invention is supported by 272 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 306 below describes the starting and ending position of this segment on each transcript.

Table 306 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1406	1460

Segment cluster HUMODCA_node_40 according to the present invention is supported by 239 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMODCA_T17. Table 307 below describes the starting and ending position of this segment on each transcript.

Table 307 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMODCA_T17	1634	1727

Variant protein alignment to the previously known protein:

Sequence name: /tmp/y03EwE6i01/dRQ5l2K6e2:DCOR_HUMAN

Sequence documentation:

Alignment of: HUMODCA_P9 x DCOR_HUMAN ..

Alignment segment 1/1:

Quality: 3056.00

Escore: 0

Matching length: 311 Total length: 311

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      30 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 79
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 200
                .         .         .         .         .
      80 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 250
                .         .         .         .         .
     130 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 179
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 300
                .         .         .         .         .
     180 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 229
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          301 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 350
                        .         .         .         .         .
          230 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 279
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          351 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 400
                        .         .         .         .         .
          280 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 329
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          401 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 450
                        .
          330 RAACASASINV                                        340
                ||||||||||||
          451 RAACASASINV                                        461
```

Sequence name: /tmp/y03EwE6i01/dRQ5l2K6e2:AAA59968

Sequence documentation:

Alignment of: HUMODCA_P9 x AAA59968 ..

Alignment segment 1/1:

Quality: 3056.00

Escore: 0

435

Matching length: 311 Total length: 311

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
          .         .         .         .         .
 30 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 79
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 40 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 89
          .         .         .         .         .
 80 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 129
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 90 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 139
          .         .         .         .         .
130 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 179
    |||||||||||||||||||||||||||||||||||||||||||||||||||
140 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 189
          .         .         .         .         .
180 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 229
    |||||||||||||||||||||||||||||||||||||||||||||||||||
190 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 239
          .         .         .         .         .
230 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 279
    |||||||||||||||||||||||||||||||||||||||||||||||||||
240 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 289
          .         .         .         .         .
280 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 329
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
        290 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 339
                       .
        330 RAACASASINV                                         340
            ||||||||||||
        340 RAACASASINV                                         350
```

Sequence name: /tmp/y03EwE6i01/dRQ5l2K6e2:AAH14562

Sequence documentation:

Alignment of: HUMODCA_P9 x AAH14562 ..

Alignment segment 1/1:

Quality: 3056.00

Escore: 0

Matching length: 311 Total length: 311

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
 30 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 79
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 86 LVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGS 135
                .         .         .         .         .
 80 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 129
    |||||||||||||||||||||||||||||||||||||||||||||||||||
136 GCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEE 185
                .         .         .         .         .
130 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 179
    |||||||||||||||||||||||||||||||||||||||||||||||||||
186 ITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQ 235
                .         .         .         .         .
180 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 229
    |||||||||||||||||||||||||||||||||||||||||||||||||||
236 TGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKY 285
                .         .         .         .         .
230 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 279
    |||||||||||||||||||||||||||||||||||||||||||||||||||
286 YSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGF 335
                .         .         .         .         .
280 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 329
    |||||||||||||||||||||||||||||||||||||||||||||||||||
336 QRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRH 385
                .
330 RAACASASINV                                         340
    ||||||||||||
386 RAACASASINV                                         396
```

DESCRIPTION FOR CLUSTER R00299

Cluster R00299 features 1 transcript(s) and 12 segment(s) of interest, the names for which are given in Tables 308 and 309, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 310.

Table 308 - Transcripts of interest

Transcript Name	Sequence ID No.
R00299_T2	33

Table 309- Segments of interest

Segment Name	Sequence ID No.
R00299_node_2	422
R00299_node_30	423
R00299_node_10	424
R00299_node_14	425
R00299_node_15	426
R00299_node_20	427
R00299_node_23	428
R00299_node_25	429
R00299_node_28	430
R00299_node_31	431
R00299_node_5	432
R00299_node_9	433

Table 310 - Proteins of interest

Protein Name	Sequence ID No.
R00299_P3	1311

These sequences are variants of the known protein Tescalcin (SwissProt accession identifier TESC_HUMAN; known also according to the synonyms TSC), SEQ ID NO: 1427, referred to herein as the previously known protein.

Protein Tescalcin is known or believed to have the following function(s): Binds calcium. The sequence for protein Tescalcin is given at the end of the application, as "Tescalcin amino acid sequence".

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: calcium binding, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster R00299 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 26 below refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 26 and Table 311. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: lung malignant tumors.

Table 311 - Normal tissue distribution

Name of Tissue	Number
bone	0
colon	0
epithelial	11
general	11
liver	0
lung	10
lymph nodes	22

440

bone marrow	31
ovary	0
pancreas	14
prostate	16
stomach	76
T cells	0
Thyroid	0

Table 312 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bone	1	6.7e-01	1	1.0	7.0e-01	1.4
colon	5.0e-02	5.3e-02	2.4e-01	2.8	2.1e-01	2.8
epithelial	7.7e-02	9.5e-02	4.0e-01	1.3	6.1e-03	1.9
general	2.3e-01	2.6e-01	5.3e-01	1.0	2.6e-04	1.9
liver	1	4.5e-01	1	1.0	6.9e-01	1.5
lung	4.9e-01	2.7e-01	6.5e-01	1.7	5.6e-04	3.8
lymph nodes	8.5e-01	8.7e-01	1	0.5	2.0e-01	1.1
bone marrow	8.6e-01	8.5e-01	1	0.5	2.3e-01	1.4
ovary	4.0e-01	4.4e-01	1	1.1	1	1.1
pancreas	7.2e-01	6.9e-01	6.7e-01	1.0	3.5e-01	1.5
prostate	8.7e-01	9.1e-01	6.7e-01	1.0	7.5e-01	0.9
stomach	6.6e-01	7.5e-01	1	0.4	6.7e-01	0.7
T cells	1	6.7e-01	1	1.0	5.2e-01	1.8
Thyroid	1.8e-01	1.8e-01	6.7e-01	1.6	6.7e-01	1.6

As noted above, cluster R00299 features 1 transcript(s), which were listed in Table 308 above. These transcript(s) encode for protein(s) which are variant(s) of protein Tescalcin. A description of each variant protein according to the present invention is now provided.

Variant protein R00299_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R00299_T2. An

alignment is given to the known protein (Tescalcin) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R00299_P3 and Q9NWT9 (SEQ ID NO:1704):

1.An isolated chimeric polypeptide encoding for R00299_P3, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV corresponding to amino acids 1 - 44 of R00299_P3, second amino acid sequence being at least 90 % homologous to SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNRNLRKGPSGLA DEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRITLEEYRNV corresponding to amino acids 74 - 191 of Q9NWT9, which also corresponds to amino acids 45 - 162 of R00299_P3, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VEELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIE TKMHVRFLNMETMALCH corresponding to amino acids 163 - 238 of R00299_P3, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV of R00299_P3.

3.An isolated polypeptide encoding for a tail of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VEELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIE TKMHVRFLNMETMALCH in R00299_P3.

Comparison report between R00299_P3 and TESC_HUMAN:

1.An isolated chimeric polypeptide encoding for R00299_P3, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV corresponding to amino acids 1 - 44 of R00299_P3, and a second amino acid sequence being at least 90 % homologous to SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNRNLRKGPSGLA DEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRITLEEYRNVVE ELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIETK MHVRFLNMETMALCH corresponding to amino acids 21 - 214 of TESC_HUMAN, which also corresponds to amino acids 45 - 238 of R00299_P3, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of R00299_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPV of R00299_P3.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Signal peptide,NN:NO) predicts that this protein has a signal peptide.

Variant protein R00299_P3 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 313, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R00299_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 313 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
120	R -> G	No
120	R -> W	No

Variant protein R00299_P3 is encoded by the following transcript(s): R00299_T2, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R00299_T2 is shown in bold; this coding portion starts at position 142 and ends at position 855. The transcript also has the following SNPs as listed in Table 314 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R00299_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 314 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
177	C -> A	Yes
499	C -> G	No
499	C -> T	No
900	G -> T	Yes
916	G ->	No
969	G ->	No
969	G -> A	No
987	A -> C	No

As noted above, cluster R00299 features 12 segment(s), which were listed in Table 309 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R00299_node_2 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 315 below describes the starting and ending position of this segment on each transcript.

Table 315 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	1	271

Segment cluster R00299_node_30 according to the present invention is supported by 75 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 316 below describes the starting and ending position of this segment on each transcript.

Table 316 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	790	961

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R00299_node_10 according to the present invention is supported by 46 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 317 below describes the starting and ending position of this segment on each transcript.

Table 317 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

R00299_T2	346	422

Segment cluster R00299_node_14 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 318 below describes the starting and ending position of this segment on each transcript.

Table 318 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	423	537

Segment cluster R00299_node_15 according to the present invention can be found in the following transcript(s): R00299_T2. Table 319 below describes the starting and ending position of this segment on each transcript.

Table 319 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	538	562

Segment cluster R00299_node_20 according to the present invention is supported by 66 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 320 below describes the starting and ending position of this segment on each transcript.

Table 320 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	563	624

Segment cluster R00299_node_23 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 321 below describes the starting and ending position of this segment on each transcript.

Table 321 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	625	732

Segment cluster R00299_node_25 according to the present invention is supported by 62 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 322 below describes the starting and ending position of this segment on each transcript.

Table 322 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	733	780

Segment cluster R00299_node_28 according to the present invention can be found in the following transcript(s): R00299_T2. Table 323 below describes the starting and ending position of this segment on each transcript.

Table 323 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	781	789

Segment cluster R00299_node_31 according to the present invention is supported by 48 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): R00299_T2. Table 324 below describes the starting and ending position of this segment on each transcript.

Table 324 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	962	1069

Segment cluster R00299_node_5 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R00299_T2. Table 325 below describes the starting and ending position of this segment on each transcript.

Table 325 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	272	341

Segment cluster R00299_node_9 according to the present invention can be found in the following transcript(s): R00299_T2. Table 326 below describes the starting and ending position of this segment on each transcript.

Table 326 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R00299_T2	342	345

Microarray (chip) data is also available for this gene as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotide was found to hit this segment (with regard to lung cancer), shown in Table 327.

Table 327 - Oligonucleotide related to this gene

Oligonucleotide name	Overexpressed in cancers	Chip reference
R00299_0_8_0	lung cancer	Lung

Variant protein alignment to the previously known protein:

Sequence name: /tmp/OleVDhrKQ0/EjblgLomjM:Q9NWT9

Sequence documentation:

Alignment of: R00299_P3 x Q9NWT9 ..

Alignment segment 1/1:

Quality: 1162.00

Escore: 0

Matching length: 118 Total length: 118

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                    .         .         .         .         .
        45 SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNR 94
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
        74 SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNR 123
```

```
                .         .         .         .         .
 95 NLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFH 144
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
124 NLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFH 173

                .
145 MYDSDSDGRITLEEYRNV                                  162
    ||||||||||||||||||
174 MYDSDSDGRITLEEYRNV                                  191
```

Sequence name: /tmp/OleVDhrKQ0/EjblgLomjM:TESC_HUMAN

Sequence documentation:

Alignment of: R00299_P3 x TESC_HUMAN ..

Alignment segment 1/1:

Quality: 1920.00

Escore: 0

Matching length: 194 Total length: 194

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

```
Alignment:
                    .         .         .         .         .
         45 SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNR 94
            |||||||||||||||||||||||||||||||||||||||||||||||||||
         21 SSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNR 70
                    .         .         .         .         .
         95 NLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFH 144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
         71 NLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFH 120
                    .         .         .         .         .
        145 MYDSDSDGRITLEEYRNVVEELLSGNPHIEKESARSIADGAMMEAASVCM 194
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
        121 MYDSDSDGRITLEEYRNVVEELLSGNPHIEKESARSIADGAMMEAASVCM 170
                    .         .         .         .
        195 GQMEPDQVYEGITFEDFLKIWQGIDIETKMHVRFLNMETMALCH       238
            ||||||||||||||||||||||||||||||||||||||||||||||||
        171 GQMEPDQVYEGITFEDFLKIWQGIDIETKMHVRFLNMETMALCH       214
```

DESCRIPTION FOR CLUSTER W60282

Cluster W60282 features 1 transcript(s) and 6 segment(s) of interest, the names for which are given in Tables 328 and 329, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 330.

Table 328 - Transcripts of interest

Transcript Name	Sequence ID No.
W60282_PEA_1_T11	34

Table 329 - Segments of interest

Segment Name	Sequence ID No.
W60282_PEA_1_node_10	434

W60282_PEA_1_node_18	435
W60282_PEA_1_node_22	436
W60282_PEA_1_node_5	437
W60282_PEA_1_node_21	438
W60282_PEA_1_node_8	439

Table 330 - Proteins of interest

Protein Name	Sequence ID No.
W60282_PEA_1_P14	1312

These sequences are variants of the known protein Kallikrein 11 precursor (SwissProt accession identifier KLKB_HUMAN; known also according to the synonyms EC 3.4.21.-; Hippostasin; Trypsin-like protease), SEQ ID NO: 1428, referred to herein as the previously known protein.

Protein Kallikrein 11 precursor is known or believed to have the following function(s): Possible multifunctional protease. Efficiently cleaves bz-Phe-Arg-4-methylcoumaryl-7-amide, a kallikrein substrate, and weakly cleaves other substrates for kallikrein and trypsin. The sequence for protein Kallikrein 11 precursor is given at the end of the application, as "Kallikrein 11 precursor amino acid sequence". Protein Kallikrein 11 precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: proteolysis and peptidolysis, which are annotation(s) related to Biological Process; and chymotrypsin; trypsin; serine-type peptidase; hydrolase, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster W60282 features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Kallikrein 11

precursor. A description of each variant protein according to the present invention is now provided.

Variant protein W60282_PEA_1_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) W60282_PEA_1_T11. An alignment is given to the known protein (Kallikrein 11 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between W60282_PEA_1_P14 and Q8IXD7 (SEQ ID NO:1705):

1.An isolated chimeric polypeptide encoding for W60282_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTA AHCLKP corresponding to amino acids 1 - 66 of Q8IXD7, which also corresponds to amino acids 1 - 66 of W60282_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TPASHLAMRQHHHH corresponding to amino acids 67 - 80 of W60282_PEA_1_P14, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of W60282_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TPASHLAMRQHHHH in W60282_PEA_1_P14.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein W60282_PEA_1_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 331, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein W60282_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 331 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
17	G -> E	Yes
41	E -> K	No

Variant protein W60282_PEA_1_P14 is encoded by the following transcript(s): W60282_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript W60282_PEA_1_T11 is shown in bold; this coding portion starts at position 705 and ends at position 944. The transcript also has the following SNPs as listed in Table 332 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein W60282_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 332- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
219	A -> G	Yes
702	G -> A	Yes
754	G -> A	Yes
825	G -> A	No
1289	A -> G	Yes

As noted above, cluster W60282 features 6 segment(s), which were listed in Table 329 above and for which the sequence(s) are given at the end of the application. These segment(s)

are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster W60282_PEA_1_node_10 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 333 below describes the starting and ending position of this segment on each transcript.

Table 333 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	745	901

Segment cluster W60282_PEA_1_node_18 according to the present invention is supported by 49 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 334 below describes the starting and ending position of this segment on each transcript.

Table 334 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	902	1038

Segment cluster W60282_PEA_1_node_22 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 335 below describes the starting and ending position of this segment on each transcript.

Table 335- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	1072	1507

Segment cluster W60282_PEA_1_node_5 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 336 below describes the starting and ending position of this segment on each transcript.

Table 336- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	1	669

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster W60282_PEA_1_node_21 according to the present invention is supported by 48 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 337 below describes the starting and ending position of this segment on each transcript.

Table 337 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	1039	1071

Segment cluster W60282_PEA_1_node_8 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): W60282_PEA_1_T11. Table 338 below describes the starting and ending position of this segment on each transcript.

Table 338 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
W60282_PEA_1_T11	670	744

Variant protein alignment to the previously known protein:

Sequence name: /tmp/rL7Wdc5hYg/eLOAfKIgqD:KLKB_HUMAN

Sequence documentation:

Alignment of: W60282_PEA_1_P14 x KLKB_HUMAN ..

Alignment segment 1/1:

Quality: 645.00

Escore: 0

Matching length: 72 Total length: 72

Matching Percent Similarity: 94.44 Matching Percent Identity: 94.44

Total Percent Similarity: 94.44 Total Percent Identity: 94.44

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGAT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGAT 50
```

```
                .         .
51 LIAPRWLLTAAHCLKPTPASHL                                    72
   ||||||||||||||||    ||
51 LIAPRWLLTAAHCLKPRYIVHL                                    72
```

Sequence name: /tmp/rL7Wdc5hYg/eLOAfKIgqD:Q8IXD7

Sequence documentation:

Alignment of: W60282_PEA_1_P14 x Q8IXD7 ..

Alignment segment 1/1:

Quality: 642.00
Escore: 0
Matching length: 66 Total length: 66
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
1 MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGAT 50
```

```
                                                458
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
          1 MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGAT 50

                                  .
         51 LIAPRWLLTAAHCLKP                                    66
5                 ||||||||||||||||
         51 LIAPRWLLTAAHCLKP                                    66
```

DESCRIPTION FOR CLUSTER Z41644

Cluster Z41644 features 1 transcript(s) and 21 segment(s) of interest, the names for which are given in Tables 339 and 340, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 341.

Table 339 - Transcripts of interest

Transcript Name	Sequence ID No.
Z41644_PEA_1_T5	35

Table 340 - Segments of interest

Segment Name	Sequence ID No.
Z41644_PEA_1_node_0	440
Z41644_PEA_1_node_11	441
Z41644_PEA_1_node_12	442
Z41644_PEA_1_node_15	443
Z41644_PEA_1_node_20	444
Z41644_PEA_1_node_24	445
Z41644_PEA_1_node_1	446
Z41644_PEA_1_node_10	447
Z41644_PEA_1_node_13	448
Z41644_PEA_1_node_16	449
Z41644_PEA_1_node_17	450

Z41644_PEA_1_node_19	451
Z41644_PEA_1_node_2	452
Z41644_PEA_1_node_21	453
Z41644_PEA_1_node_22	454
Z41644_PEA_1_node_23	455
Z41644_PEA_1_node_25	456
Z41644_PEA_1_node_3	457
Z41644_PEA_1_node_4	458
Z41644_PEA_1_node_6	459
Z41644_PEA_1_node_9	460

Table 341 - Proteins of interest

Protein Name	Sequence ID No.
Z41644_PEA_1_P10	1313

These sequences are variants of the known protein Small inducible cytokine B14 precursor (SwissProt accession identifier SZ14_HUMAN; known also according to the synonyms CXCL14; Chemokine BRAK), SEQ ID NO:1429, referred to herein as the previously known protein.

The sequence for protein Small inducible cytokine B14 precursor is given at the end of the application, as "Small inducible cytokine B14 precursor amino acid sequence". Protein Small inducible cytokine B14 precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: chemotaxis; signal transduction; cell-cell signaling, which are annotation(s) related to Biological Process; and chemokine, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster Z41644 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 27 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 27 and Table 342. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: lung malignant tumors, breast malignant tumors and pancreas carcinoma.

Table 342 - Normal tissue distribution

Name of Tissue	Number
bone	45
brain	62
colon	327
epithelial	179
general	104
head and neck	10
kidney	219
lung	6
lymph nodes	37
breast	87
bone marrow	0
muscle	20
ovary	36
pancreas	0
prostate	78
skin	591

stomach	109
Thyroid	386
uterus	218

Table 343 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bone	4.9e-01	8.5e-01	1.8e-01	1.9	5.3e-01	1.0
brain	6.7e-01	8.0e-01	9.1e-01	0.6	9.9e-01	0.4
colon	6.4e-01	7.7e-01	9.7e-01	0.4	1	0.3
epithelial	4.1e-01	9.4e-01	9.6e-01	0.7	1	0.4
general	1.5e-01	9.4e-01	1.8e-01	1.0	1	0.5
head and neck	1.9e-01	3.3e-01	4.6e-01	2.8	7.5e-01	1.5
kidney	7.7e-01	8.2e-01	7.0e-01	0.7	9.5e-01	0.5
lung	2.2e-01	5.0e-01	1.3e-04	8.7	8.1e-03	4.1
lymph nodes	6.3e-01	8.7e-01	6.3e-01	1.2	9.2e-01	0.6
breast	4.0e-01	6.5e-01	3.9e-04	3.5	2.9e-02	1.9
bone marrow	1	6.7e-01	1	1.0	5.3e-01	1.9
muscle	5.2e-01	6.1e-01	2.7e-01	3.2	6.3e-01	1.2
ovary	6.7e-01	7.1e-01	7.6e-01	1.0	8.6e-01	0.8
pancreas	2.2e-02	2.3e-02	5.7e-03	7.8	1.6e-03	8.2
prostate	8.8e-01	9.0e-01	8.3e-01	0.6	9.3e-01	0.5
skin	5.9e-01	6.9e-01	2.3e-01	0.3	1	0.0
stomach	6.1e-01	8.9e-01	8.1e-01	0.7	9.9e-01	0.4
Thyroid	7.0e-01	7.0e-01	9.9e-01	0.4	9.9e-01	0.4
uterus	5.3e-01	8.2e-01	9.5e-01	0.5	1	0.3

As noted above, cluster Z41644 features 1 transcript(s), which were listed in Table 339 above. These transcript(s) encode for protein(s) which are variant(s) of protein Small inducible cytokine B14 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein Z41644_PEA_1_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z41644_PEA_1_T5. An alignment is given to the known protein (Small inducible cytokine B14 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z41644_PEA_1_P10 and SZ14_HUMAN:

1.An isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 1 - 95 of SZ14_HUMAN, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to amino acids 96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

Comparison report between Z41644_PEA_1_P10 and Q9NS21 (SEQ ID NO:1706):

1.An isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 13 - 107 of Q9NS21, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to amino acids

96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

Comparison report between Z41644_PEA_1_P10 and AAQ89265 (SEQ ID NO:781):

1.An isolated chimeric polypeptide encoding for Z41644_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVII TTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR corresponding to amino acids 13 - 107 of AAQ89265, which also corresponds to amino acids 1 - 95 of Z41644_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI corresponding to amino acids 96 - 123 of Z41644_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z41644_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homolo gous to the sequence YAPPLLTFLPTRPSCGSQDGKGPPHQVI in Z41644_PEA_1_P10.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z41644_PEA_1_P10 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 344, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether

the SNP is known or not; the presence of known SNPs in variant protein Z41644_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 344 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
32	P -> H	Yes
64	S ->	No
80	T -> A	No
80	T -> P	No

Variant protein Z41644_PEA_1_P10 is encoded by the following transcript(s): Z41644_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z41644_PEA_1_T5 is shown in bold; this coding portion starts at position 744 and ends at position 1112. The transcript also has the following SNPs as listed in Table 345 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z41644_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 345 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
102	A -> G	Yes
572	C ->	No
3707	C -> T	Yes
3735	C -> T	Yes
4079	G -> A	No
4123	G -> A	Yes
4233	A -> G	Yes

4328	C ->	No
4350	A -> G	Yes
4376	G -> A	Yes
4390	A -> G	Yes
4619	G -> T	Yes
838	C -> A	Yes
4754	C -> T	No
4757	C -> A	No
4794	T -> G	No
4827	G ->	No
934	C ->	No
981	A -> C	No
981	A -> G	No
1817	A -> C	Yes
2546	T ->	No
2684	T -> A	No
2885	T -> C	Yes

As noted above, cluster Z41644 features 21 segment(s), which were listed in Table 340 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster Z41644_PEA_1_node_0 according to the present invention is supported by 53 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 346 below describes the starting and ending position of this segment on each transcript.

Table 346 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	1	616

Segment cluster Z41644_PEA_1_node_11 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 347 below describes the starting and ending position of this segment on each transcript.

Table 347 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	1028	2089

Segment cluster Z41644_PEA_1_node_12 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 348 below describes the starting and ending position of this segment on each transcript.

Table 348 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	2090	2350

Segment cluster Z41644_PEA_1_node_15 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 349 below describes the starting and ending position of this segment on each transcript.

Table 349 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	2368	3728

Segment cluster Z41644_PEA_1_node_20 according to the present invention is supported by 260 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 350 below describes the starting and ending position of this segment on each transcript.

Table 350 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	3938	4506

Segment cluster Z41644_PEA_1_node_24 according to the present invention is supported by 185 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 351 below describes the starting and ending position of this segment on each transcript.

Table 351 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	4637	4799

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster Z41644_PEA_1_node_1 according to the present invention is supported by 53 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 352 below describes the starting and ending position of this segment on each transcript.

Table 352 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	617	697

Segment cluster Z41644_PEA_1_node_10 according to the present invention is supported by 138 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 353 below describes the starting and ending position of this segment on each transcript.

Table 353 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	972	1027

Segment cluster Z41644_PEA_1_node_13 according to the present invention can be found in the following transcript(s): Z41644_PEA_1_T5. Table 354 below describes the starting and ending position of this segment on each transcript.

Table 354 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	2351	2367

Segment cluster Z41644_PEA_1_node_16 according to the present invention is supported by 152 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 355 below describes the starting and ending position of this segment on each transcript.

Table 355 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	3729	3809

Segment cluster Z41644_PEA_1_node_17 according to the present invention can be found in the following transcript(s): Z41644_PEA_1_T5. Table 356 below describes the starting and ending position of this segment on each transcript.

Table 356 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	3810	3829

Segment cluster Z41644_PEA_1_node_19 according to the present invention is supported by 112 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 357 below describes the starting and ending position of this segment on each transcript.

Table 357 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	3830	3937

Segment cluster Z41644_PEA_1_node_2 according to the present invention is supported by 58 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 358 below describes the starting and ending position of this segment on each transcript.

Table 358 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	698	737

Segment cluster Z41644_PEA_1_node_21 according to the present invention can be found in the following transcript(s): Z41644_PEA_1_T5. Table 359 below describes the starting and ending position of this segment on each transcript.

Table 359 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

Z41644_PEA_1_T5	4507	4529

Segment cluster Z41644_PEA_1_node_22 according to the present invention is supported by 164 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 360 below describes the starting and ending position of this segment on each transcript.

Table 360 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	4530	4582

Segment cluster Z41644_PEA_1_node_23 according to the present invention is supported by 169 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 361 below describes the starting and ending position of this segment on each transcript.

Table 361 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	4583	4636

Segment cluster Z41644_PEA_1_node_25 according to the present invention is supported by 138 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 362 below describes the starting and ending position of this segment on each transcript.

Table 362 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	4800	4902

Segment cluster Z41644_PEA_1_node_3 according to the present invention is supported by 75 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 363 below describes the starting and ending position of this segment on each transcript.

Table 363 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	738	773

Segment cluster Z41644_PEA_1_node_4 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 364 below describes the starting and ending position of this segment on each transcript.

Table 364 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	774	807

Segment cluster Z41644_PEA_1_node_6 according to the present invention is supported by 101 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 365 below describes the starting and ending position of this segment on each transcript.

Table 365 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	808	913

Segment cluster Z41644_PEA_1_node_9 according to the present invention is supported by 134 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z41644_PEA_1_T5. Table 366 below describes the starting and ending position of this segment on each transcript.

Table 366 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z41644_PEA_1_T5	914	971

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/p5SSvhT9Xp/HQeIMsUrfm:SZ14_HUMAN

Sequence documentation:

Alignment of: Z41644_PEA_1_P10 x SZ14_HUMAN    ..

Alignment segment 1/1:

                      Quality:  953.00
Escore:       0
             Matching length:      95                 Total
length:      95
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                         Gaps:       0
```

```
Alignment:

                    .         .         .         .         .
          1 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 50
            ||||||||||||||||||||||||||||||||||||||||||||||||||
          1 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 50
                    .         .         .         .
         51 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR       95
            |||||||||||||||||||||||||||||||||||||||||||||
         51 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR       95
```

Sequence name: /tmp/p5SSvhT9Xp/HQeIMsUrfm:Q9NS21

Sequence documentation:

Alignment of: Z41644_PEA_1_P10 x Q9NS21 ..

Alignment segment 1/1:

Quality: 957.00

Escore: 0

Matching length: 96 Total length: 96

Matching Percent Similarity: 100.00 Matching Percent Identity: 98.96

Total Percent Similarity: 100.00 Total Percent Identity: 98.96

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     13 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 62
                .         .         .         .
     51 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRRY     96
        ||||||||||||||||||||||||||||||||||||||||||||||||:
     63 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRRF     108
```

Sequence name: /tmp/p5SSvhT9Xp/HQeIMsUrfm:AAQ89265

Sequence documentation:

Alignment of: Z41644_PEA_1_P10 x AAQ89265 ..

Alignment segment 1/1:

Quality: 953.00

Escore: 0

Matching length: 95 Total length: 95

475

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       13 MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH 62
            .   .         .         .         .
       51 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR      95
          |||||||||||||||||||||||||||||||||||||||||||||
       63 CEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRR      107
```

DESCRIPTION FOR CLUSTER Z44808

Cluster Z44808 features 5 transcript(s) and 21 segment(s) of interest, the names for which are given in Tables 367 and 368, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 369.

Table 367 - Transcripts of interest

Transcript Name	Sequence ID No.
Z44808_PEA_1_T11	36
Z44808_PEA_1_T4	37
Z44808_PEA_1_T5	38
Z44808_PEA_1_T8	39
Z44808_PEA_1_T9	40

Table 368 - Segments of interest

Segment Name	Sequence ID No.
Z44808_PEA_1_node_0	461
Z44808_PEA_1_node_16	462
Z44808_PEA_1_node_2	463
Z44808_PEA_1_node_24	464
Z44808_PEA_1_node_32	465
Z44808_PEA_1_node_33	466
Z44808_PEA_1_node_36	467
Z44808_PEA_1_node_37	468
Z44808_PEA_1_node_41	469
Z44808_PEA_1_node_11	470
Z44808_PEA_1_node_13	471
Z44808_PEA_1_node_18	472
Z44808_PEA_1_node_22	473
Z44808_PEA_1_node_26	474
Z44808_PEA_1_node_30	475
Z44808_PEA_1_node_34	476
Z44808_PEA_1_node_35	477
Z44808_PEA_1_node_39	478
Z44808_PEA_1_node_4	479
Z44808_PEA_1_node_6	480
Z44808_PEA_1_node_8	481

Table 369 - Proteins of interest

Protein Name	Sequence ID No.
Z44808_PEA_1_P5	1314
Z44808_PEA_1_P6	1315
Z44808_PEA_1_P7	1316
Z44808_PEA_1_P11	1317

These sequences are variants of the known protein SPARC related modular calcium-binding protein 2 precursor (SwissProt accession identifier SMO2_HUMAN; known also according to the synonyms Secreted modular calcium-binding protein 2; SMOC-2; Smooth muscle-associated protein 2; SMAP-2; MSTP117), SEQ ID NO: 1430, referred to herein as the previously known protein.

Protein SPARC related modular calcium-binding protein 2 precursor is known or believed to have the following function(s): calcium binding. The sequence for protein SPARC related modular calcium-binding protein 2 precursor is given at the end of the application, as "SPARC related modular calcium-binding protein 2 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 370.

Table 370 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
169 - 170	KT -> TR
212	S -> P
429 - 446	TPRGHAESTSNRQPRKQG -> RSKRNL
434	A -> V
439	N -> Y

Protein SPARC related modular calcium-binding protein 2 precursor localization is believed to be Secreted.

Cluster Z44808 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 28 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 28 and Table 371. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: colorectal cancer, lung cancer and pancreas carcinoma.

Table 371 - Normal tissue distribution

Name of Tissue	Number
bladder	123
bone	304
brain	18
colon	0
epithelial	40
general	37
kidney	2
lung	0
breast	61
ovary	116
pancreas	0
prostate	128
stomach	36
uterus	195

Table 372 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bladder	6.8e-01	7.6e-01	7.7e-01	0.8	9.1e-01	0.6
bone	7.0e-01	8.8e-01	9.9e-01	0.3	1	0.2
brain	6.8e-01	7.2e-01	3.0e-02	2.6	1.7e-01	1.6
colon	9.2e-03	1.3e-02	1.2e-01	3.6	1.6e-01	3.1
epithelial	2.1e-02	4.0e-01	1.0e-04	1.9	2.7e-01	1.0
general	2.6e-02	7.2e-01	4.9e-07	1.9	3.0e-01	1.0
kidney	7.3e-01	8.1e-01	1	1.0	1	1.0

lung	4.0e-03	1.8e-02	8.0e-04	12.2	2.1e-02	6.0
breast	4.8e-01	6.1e-01	9.8e-02	2.0	3.9e-01	1.2
ovary	8.1e-01	8.3e-01	9.1e-01	0.6	9.7e-01	0.5
pancreas	1.2e-01	2.1e-01	1.0e-03	6.5	5.9e-03	4.6
prostate	8.4e-01	8.9e-01	9.0e-01	0.6	9.8e-01	0.4
stomach	5.0e-01	8.7e-01	9.6e-04	1.5	1.9e-01	0.8
uterus	6.7e-01	7.9e-01	9.2e-01	0.5	1	0.3

As noted above, cluster Z44808 features 5 transcript(s), which were listed in Table 367 above. These transcript(s) encode for protein(s) which are variant(s) of protein SPARC related modular calcium-binding protein 2 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein Z44808_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z44808_PEA_1_T4. An alignment is given to the known protein (SPARC related modular calcium-binding protein 2 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z44808_PEA_1_P5 and SMO2_HUMAN:

1.An isolated chimeric polypeptide encoding for Z44808_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to

MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR
TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD
GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA
APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN
DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA
KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE
RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ

ELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ corresponding to amino acids 1 - 441 of SMO2_HUMAN, which also corresponds to amino acids 1 - 441 of Z44808_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DAMVVSSRPKATTHRKSRTLSRR corresponding to amino acids 442 - 464 of Z44808_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z44808_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DAMVVSSRPKATTHRKSRTLSRR in Z44808_PEA_1_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z44808_PEA_1_P5 is encoded by the following transcript(s): Z44808_PEA_1_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z44808_PEA_1_T4 is shown in bold; this coding portion starts at position 586 and ends at position 1977. The transcript also has the following SNPs as listed in Table 373 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 373 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?

549	A -> G	No
648	T -> G	No
4403	G -> T	No
4456	G -> A	Yes
4964	G -> C	Yes
1025	C ->	No
1677	T -> C	No
2691	C -> T	Yes
3900	T -> C	No
3929	G -> A	Yes
4099	G -> T	Yes
4281	T -> C	No
4319	G -> C	Yes

Variant protein Z44808_PEA_1_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z44808_PEA_1_T5. An alignment is given to the known protein (SPARC related modular calcium-binding protein 2 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z44808_PEA_1_P6 and SMO2_HUMAN:

1.An isolated chimeric polypeptide encoding for Z44808_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN

DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ ELMGCLGVAKEDGKADTKKRH corresponding to amino acids 1 - 428 of SMO2_HUMAN, which also corresponds to amino acids 1 - 428 of Z44808_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RSKRNL corresponding to amino acids 429 - 434 of Z44808_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z44808_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RSKRNL in Z44808_PEA_1_P6.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z44808_PEA_1_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 374, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 374 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
147	A ->	No

Variant protein Z44808_PEA_1_P6 is encoded by the following transcript(s): Z44808_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z44808_PEA_1_T5 is shown in bold; this coding portion starts at position 586 and ends at position 1887. The transcript also has the following SNPs as listed in Table 375 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 375 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
549	A -> G	No
648	T -> G	No
2866	G -> A	Yes
3374	G -> C	Yes
1025	C ->	No
1677	T -> C	No
2310	T -> C	No
2339	G -> A	Yes
2509	G -> T	Yes
2691	T -> C	No
2729	G -> C	Yes
2813	G -> T	No

Variant protein Z44808_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z44808_PEA_1_T9. An alignment is given to the known protein (SPARC related modular calcium-binding protein 2 precursor) at the end of the application. One or more alignments to

one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z44808_PEA_1_P7 and SMO2_HUMAN:

1.An isolated chimeric polypeptide encoding for Z44808_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAA APALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKN DNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPA KARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEE RVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQ ELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ corresponding to amino acids 1 - 441 of SMO2_HUMAN, which also corresponds to amino acids 1 - 441 of Z44808_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LLWLRGKVSFYCF corresponding to amino acids 442 - 454 of Z44808_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z44808_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LLWLRGKVSFYCF in Z44808_PEA_1_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide

prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z44808_PEA_1_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 376, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 376 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
147	A ->	No

Variant protein Z44808_PEA_1_P7 is encoded by the following transcript(s): Z44808_PEA_1_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z44808_PEA_1_T9 is shown in bold; this coding portion starts at position 586 and ends at position 1947. The transcript also has the following SNPs as listed in Table 377 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 377 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
549	A -> G	No
648	T -> G	No
1025	C ->	No
1677	T -> C	No
2169	C -> A	Yes

Variant protein Z44808_PEA_1_P11 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z44808_PEA_1_T11. The identification of this transcript was performed using a non-EST based method for identification of alternative splicing, described in the following reference: "Sorek R et al., Genome Res. (2004) 14:1617-23." An alignment is given to the known protein (SPARC related modular calcium-binding protein 2 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z44808_PEA_1_P11 and SMO2_HUMAN:

1.An isolated chimeric polypeptide encoding for Z44808_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGR TFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDD GTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCPGSVNEKLPQREGTGKT corresponding to amino acids 1 - 170 of SMO2_HUMAN, which also corresponds to amino acids 1 - 170 of Z44808_PEA_1_P11, and a second amino acid sequence being at least 90 % homologous to DIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGL YKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQ GCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLD KNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQELMGCLGVAKE DGKADTKKRHTPRGHAESTSNRQPRKQG corresponding to amino acids 188 - 446 of SMO2_HUMAN, which also corresponds to amino acids 171 - 429 of Z44808_PEA_1_P11, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of Z44808_PEA_1_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TD, having a

structure as follows: a sequence starting from any of amino acid numbers 170-x to -170; and ending at any of amino acid numbers 171+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein Z44808_PEA_1_P11 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 378, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 378 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
147	A ->	No

Variant protein Z44808_PEA_1_P11 is encoded by the following transcript(s): Z44808_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z44808_PEA_1_T11 is shown in bold; this coding portion starts at position 586 and ends at position 1872. The transcript also has the following SNPs as listed in Table 379 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z44808_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 379 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
549	A -> G	No
648	T -> G	No
2720	G -> A	Yes
3228	G -> C	Yes
1025	C ->	No
1626	T -> C	No
2164	T -> C	No
2193	G -> A	Yes
2363	G -> T	Yes
2545	T -> C	No
2583	G -> C	Yes
2667	G -> T	No

As noted above, cluster Z44808 features 21 segment(s), which were listed in Table 368 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster Z44808_PEA_1_node_0 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 380 below describes the starting and ending position of this segment on each transcript.

Table 380 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1	669
Z44808_PEA_1_T4	1	669

Z44808_PEA_1_T5	1	669
Z44808_PEA_1_T8	1	669
Z44808_PEA_1_T9	1	669

Segment cluster Z44808_PEA_1_node_16 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 381 below describes the starting and ending position of this segment on each transcript.

Table 381 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1172	1358
Z44808_PEA_1_T4	1223	1409
Z44808_PEA_1_T5	1223	1409
Z44808_PEA_1_T8	1223	1409
Z44808_PEA_1_T9	1223	1409

Segment cluster Z44808_PEA_1_node_2 according to the present invention is supported by 34 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 382 below describes the starting and ending position of this segment on each transcript.

Table 382 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	670	841
Z44808_PEA_1_T4	670	841
Z44808_PEA_1_T5	670	841

Z44808_PEA_1_T8	670	841
Z44808_PEA_1_T9	670	841

Segment cluster Z44808_PEA_1_node_24 according to the present invention is supported by 52 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 383 below describes the starting and ending position of this segment on each transcript.

Table 383 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1545	1819
Z44808_PEA_1_T4	1596	1870
Z44808_PEA_1_T5	1596	1870
Z44808_PEA_1_T8	1596	1870
Z44808_PEA_1_T9	1596	1870

Segment cluster Z44808_PEA_1_node_32 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T4 and Z44808_PEA_1_T8. Table 384 below describes the starting and ending position of this segment on each transcript.

Table 384 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T4	1909	3593
Z44808_PEA_1_T8	1909	2397

Segment cluster Z44808_PEA_1_node_33 according to the present invention is supported by 133 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4 and Z44808_PEA_1_T5. Table 385 below describes the starting and ending position of this segment on each transcript.

Table 385 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1858	2734
Z44808_PEA_1_T4	3594	4470
Z44808_PEA_1_T5	2004	2880

Segment cluster Z44808_PEA_1_node_36 according to the present invention is supported by 117 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4 and Z44808_PEA_1_T5. Table 386 below describes the starting and ending position of this segment on each transcript.

Table 386 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	2829	3080
Z44808_PEA_1_T4	4565	4816
Z44808_PEA_1_T5	2975	3226

Segment cluster Z44808_PEA_1_node_37 according to the present invention is supported by 120 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4 and Z44808_PEA_1_T5. Table 387 below describes the starting and ending position of this segment on each transcript.

Table 387 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

Z44808_PEA_1_T11	3081	3429
Z44808_PEA_1_T4	4817	5165
Z44808_PEA_1_T5	3227	3575

Segment cluster Z44808_PEA_1_node_41 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T9. Table 388 below describes the starting and ending position of this segment on each transcript.

Table 388 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T9	1974	2206

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster Z44808_PEA_1_node_11 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 389 below describes the starting and ending position of this segment on each transcript.

Table 389 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T4	1097	1147
Z44808_PEA_1_T5	1097	1147
Z44808_PEA_1_T8	1097	1147

Z44808_PEA_1_T9	1097	1147

Segment cluster Z44808_PEA_1_node_13 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 390 below describes the starting and ending position of this segment on each transcript.

Table 390 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1097	1171
Z44808_PEA_1_T4	1148	1222
Z44808_PEA_1_T5	1148	1222
Z44808_PEA_1_T8	1148	1222
Z44808_PEA_1_T9	1148	1222

Segment cluster Z44808_PEA_1_node_18 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 391 below describes the starting and ending position of this segment on each transcript.

Table 391 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1359	1441
Z44808_PEA_1_T4	1410	1492
Z44808_PEA_1_T5	1410	1492
Z44808_PEA_1_T8	1410	1492
Z44808_PEA_1_T9	1410	1492

Segment cluster Z44808_PEA_1_node_22 according to the present invention is supported by 33 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 392 below describes the starting and ending position of this segment on each transcript.

Table 392 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1442	1544
Z44808_PEA_1_T4	1493	1595
Z44808_PEA_1_T5	1493	1595
Z44808_PEA_1_T8	1493	1595
Z44808_PEA_1_T9	1493	1595

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (with regard to lung cancer), shown in Table 393.

Table 393 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
Z44808_0_8_0	Lung squamous cell carcinoma	LUN

Segment cluster Z44808_PEA_1_node_26 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T5. Table 394 below describes the starting and ending position of this segment on each transcript.

Table 394 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T5	1871	1965

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (with regard to lung cancer), shown in Table 395.

Table 395 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
Z44808_0_0_72347	Lung small cell cancer	LUN

Segment cluster Z44808_PEA_1_node_30 according to the present invention is supported by 44 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 396 below describes the starting and ending position of this segment on each transcript.

Table 396 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1820	1857
Z44808_PEA_1_T4	1871	1908
Z44808_PEA_1_T5	1966	2003
Z44808_PEA_1_T8	1871	1908
Z44808_PEA_1_T9	1871	1908

Segment cluster Z44808_PEA_1_node_34 according to the present invention is supported by 70 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4 and Z44808_PEA_1_T5. Table 397 below describes the starting and ending position of this segment on each transcript.

Table 397 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	2735	2809
Z44808_PEA_1_T4	4471	4545
Z44808_PEA_1_T5	2881	2955

Segment cluster Z44808_PEA_1_node_35 according to the present invention can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4 and Z44808_PEA_1_T5. Table 398 below describes the starting and ending position of this segment on each transcript.

Table 398 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	2810	2828
Z44808_PEA_1_T4	4546	4564
Z44808_PEA_1_T5	2956	2974

Segment cluster Z44808_PEA_1_node_39 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T9. Table 399 below describes the starting and ending position of this segment on each transcript.

Table 399 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T9	1909	1973

Segment cluster Z44808_PEA_1_node_4 according to the present invention is supported by 33 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 400 below describes the starting and ending position of this segment on each transcript.

Table 400 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	842	948
Z44808_PEA_1_T4	842	948
Z44808_PEA_1_T5	842	948
Z44808_PEA_1_T8	842	948
Z44808_PEA_1_T9	842	948

Segment cluster Z44808_PEA_1_node_6 according to the present invention is supported by 30 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 401 below describes the starting and ending position of this segment on each transcript.

Table 401 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	949	1048
Z44808_PEA_1_T4	949	1048
Z44808_PEA_1_T5	949	1048
Z44808_PEA_1_T8	949	1048
Z44808_PEA_1_T9	949	1048

Segment cluster Z44808_PEA_1_node_8 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z44808_PEA_1_T11, Z44808_PEA_1_T4, Z44808_PEA_1_T5, Z44808_PEA_1_T8 and Z44808_PEA_1_T9. Table 402 below describes the starting and ending position of this segment on each transcript.

Table 402 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z44808_PEA_1_T11	1049	1096
Z44808_PEA_1_T4	1049	1096
Z44808_PEA_1_T5	1049	1096
Z44808_PEA_1_T8	1049	1096
Z44808_PEA_1_T9	1049	1096

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/vUqLu6eAVZ/K3JDuPvaLo:SMO2_HUMAN

Sequence documentation:

Alignment of: Z44808_PEA_1_P5 x SMO2_HUMAN    ..

Alignment segment 1/1:

                        Quality: 4440.00
Escore:         0
```

Matching length: 441 Total length: 441

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
                .         .         .         .         .
     51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
                .         .         .         .         .
    101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
                .         .         .         .         .
    151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
                .         .         .         .         .
    201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
                .         .         .         .         .
    251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
```

```
                                500
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
          251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
                         .         .         .         .         .
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                         .         .         .         .         .
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                         .         .         .         .
          401 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ           441
                    ||||||||||||||||||||||||||||||||||||||||
          401 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ           441
```

Sequence name: /tmp/QSUNfTsJ5y/kLOw5Vb6SD:SMO2_HUMAN

Sequence documentation:

Alignment of: Z44808_PEA_1_P6 x SMO2_HUMAN ..

Alignment segment 1/1:

Quality: 4310.00

Escore: 0

501

Matching length: 428 Total length: 428

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
                .         .         .         .         .
 51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
                .         .         .         .         .
101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
                .         .         .         .         .
151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
                .         .         .         .         .
201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
                .         .         .         .         .
251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
                        .         .         .         .         .
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                        .         .         .         .         .
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                        .         .
          401 DKSISVQELMGCLGVAKEDGKADTKKRH                       428
                ||||||||||||||||||||||||||||
          401 DKSISVQELMGCLGVAKEDGKADTKKRH                       428
```

Sequence name: /tmp/MZVdR4PVdM/5uN8RwViJ1:SMO2_HUMAN

Sequence documentation:

Alignment of: Z44808_PEA_1_P7 x SMO2_HUMAN ..

Alignment segment 1/1:

Quality: 4440.00

Escore: 0

Matching length: 441 Total length: 441

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50

                .         .         .         .         .
     51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100

                .         .         .         .         .
    101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150

                .         .         .         .         .
    151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200

                .         .         .         .         .
    201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250

                .         .         .         .         .
    251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
                         .         .         .         .         .
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                         .         .         .         .         .
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                         .         .         .         .
          401 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ           441
                ||||||||||||||||||||||||||||||||||||||||
          401 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQ           441
```

Sequence name: /tmp/3fGVxqLloe/J5mQduAd0F:SMO2_HUMAN

Sequence documentation:

Alignment of: Z44808_PEA_1_P11 x SMO2_HUMAN ..

Alignment segment 1/1:

Quality: 4228.00

Escore: 0

505

Matching length: 429 Total length: 446

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 96.19 Total Percent Identity: 96.19

Gaps: 1

Alignment:

```
          .         .         .         .         .
  1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
    |||||||||||||||||||||||||||||||||||||||||||||||||
  1 MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQK 50
          .         .         .         .         .
 51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
    |||||||||||||||||||||||||||||||||||||||||||||||||
 51 PLCASDGRTFLSRCEFQRAKCKDPQLEIAYRGNCKDVSRCVAERKYTQEQ 100
          .         .         .         .         .
101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
    |||||||||||||||||||||||||||||||||||||||||||||||||
101 ARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKT 150
          .         .         .         .         .
151 PRCPGSVNEKLPQREGTGKT.................DIASRYPTLWTEQ 183
    ||||||||||||||||||                  ||||||||||||
151 PRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQ 200
          .         .         .         .         .
184 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 233
    |||||||||||||||||||||||||||||||||||||||||||||||||
201 VKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQ 250
          .         .         .         .         .
234 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 283
```

506

```
                                                                                
          251 CHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQ 300
                    .         .         .         .         .
          284 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 333
                                                                                
          301 LQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER 350
                    .         .         .         .         .
          334 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 383
                                                                                
          351 VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNN 400
                    .         .         .         .
          384 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQPRKQG      429
                                                                                
          401 DKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQPRKQG      446
```

Expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor Z44808 transcripts which are detectable by amplicon as depicted in sequence name Z44808junc8-11 in normal and cancerous lung tissues

Expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor (Secreted modular calcium-binding protein 2) (SMOC-2) (Smooth muscle-associated protein 2) transcripts detectable by or according to junc8-11, Z44808 junc8-11 amplicon (SEQ ID NO: 1651) and Z44808junc8-11F (SEQ ID NO: 1649) and Z44808junc8-11R (SEQ ID NO: 1650) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon– HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping

genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 29 is a histogram showing over expression of the above-indicated SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 29, the expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor transcripts detectable by the above amplicon in several cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 2 out of 15 adenocarcinoma samples and in 3 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: *Z44808junc8-11F* forward primer; and *Z44808junc8-11 R* reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: *Z44808junc8-11*
Forward primer (SEQ ID NO: 1649): GAAGGCACAGGAAAAACAGATATTG
Reverse primer (SEQ ID NO: 1650): TGGTGCTCTTGGTCACAGGAT
Amplicon (SEQ ID NO: 1651):
GAAGGCACAGGAAAAACAGATATTGCATCACGTTACCCTACCCTTTGGACTGAACA
GGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTGTGACCAAG
AGCACCA

Expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor (Secreted modular calcium-binding protein 2) (SMOC-2) (Smooth muscle-associated protein 2) Z44808 transcripts which are detectable by amplicon as depicted in sequence name Z44808 junc8-11 in different normal tissues

Expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor (Secreted modular calcium-binding protein 2) (SMOC-2) (Smooth muscle-associated protein 2) transcripts detectable by or according to *Z44808 junc8-11* amplicon (SEQ ID NO: 1651) and primers: *Z44808 junc8-11F* (SEQ ID NO: 1649) and *Z44808 junc8-11R* (SEQ ID NO: 1650) was measured by real time PCR. In parallel the expression of four housekeeping genes – RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20, Table 3), to obtain a value of relative expression of each sample relative to median of the ovary samples.

Primers:

Forward primer (SEQ ID NO: 1649): GAAGGCACAGGAAAAACAGATATTG

Reverse primer (SEQ ID NO: 1650): TGGTGCTCTTGGTCACAGGAT

Amplicon (SEQ ID NO: 1651):

GAAGGCACAGGAAAAACAGATATTGCATCACGTTACCCTACCCTTTGGACTGAACA GGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTGTGACCAAG AGCACCA

The results are demonstrated in Figure 18, showing the expression of SMO2_HUMAN SPARC related modular calcium-binding protein 2 precursor (Secreted modular calcium-binding protein 2) (SMOC-2) (Smooth muscle-associated protein 2) Z44808 transcripts which are detectable by amplicon as depicted in sequence name Z44808 junc8-11 in different normal tissues.

DESCRIPTION FOR CLUSTER AA161187

Cluster AA161187 features 7 transcript(s) and 20 segment(s) of interest, the names for which are given in Tables 403 and 404, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 405.

Table 403 - Transcripts of interest

Transcript Name	Sequence ID No.
AA161187_T0	41
AA161187_T7	42
AA161187_T15	43
AA161187_T16	44
AA161187_T20	45
AA161187_T21	46
AA161187_T22	47

Table 404 - Segments of interest

Segment Name	Sequence ID No.
AA161187_node_0	482
AA161187_node_6	483
AA161187_node_14	484
AA161187_node_16	485
AA161187_node_25	486
AA161187_node_26	487
AA161187_node_28	488

AA161187_node_4	489
AA161187_node_7	490
AA161187_node_8	491
AA161187_node_9	492
AA161187_node_10	493
AA161187_node_12	494
AA161187_node_13	495
AA161187_node_19	496
AA161187_node_20	497
AA161187_node_21	498
AA161187_node_22	499
AA161187_node_23	500
AA161187_node_24	501

Table 405 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
AA161187_P1	1318	AA161187_T0
AA161187_P6	1319	AA161187_T7
AA161187_P13	1320	AA161187_T15
AA161187_P14	1321	AA161187_T16
AA161187_P18	1322	AA161187_T20
AA161187_P19	1323	AA161187_T21

These sequences are variants of the known protein Testisin precursor (SwissProt accession identifier TEST_HUMAN; known also according to the synonyms EC 3.4.21.-; Eosinophil serine protease 1; ESP- 1; UNQ266/PRO303), SEQ ID NO: 1431, referred to herein as the previously known protein.

Protein Testisin precursor is known or believed to have the following function(s): Could regulate proteolytic events associated with testicular germ cell maturation. The sequence for protein Testisin precursor is given at the end of the application, as "Testisin precursor amino

acid sequence". Protein Testisin precursor localization is believed to be attached to the membrane by a GPI-anchor.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: serine-type peptidase, which are annotation(s) related to Molecular Function; and membrane fraction; cytoplasm; plasma membrane, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster AA161187 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the left hand column of the table and the numbers on the y-axis of figure 30 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 30 and Table 406. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: brain malignant tumors, epithelial malignant tumors and a mixture of malignant tumors from different tissues.

Table 406 - Normal tissue distribution

Name of Tissue	Number
bone	0
brain	1
colon	0
epithelial	0
general	0
lung	0

breast	0
bone marrow	0
ovary	0
pancreas	0
prostate	4
stomach	0
uterus	0

Table 407 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bone	1	6.7e-01	1	1.0	3.4e-01	1.9
brain	9.8e-01	6.0e-01	1	0.7	3.8e-03	3.6
colon	4.4e-01	5.0e-01	7.0e-01	1.5	7.7e-01	1.3
epithelial	1.3e-02	2.6e-03	1.7e-03	8.4	2.4e-04	7.9
general	1.6e-03	1.9e-05	1.9e-05	12.1	2.9e-10	15.6
lung	5.0e-01	6.3e-01	1.7e-01	3.9	3.8e-01	2.2
breast	1	6.7e-01	1	1.0	8.2e-01	1.2
bone marrow	1	4.2e-01	1	1.0	1.5e-01	2.9
ovary	6.2e-01	6.5e-01	4.7e-01	1.9	5.9e-01	1.6
pancreas	1	4.4e-01	1	1.0	2.8e-01	2.8
prostate	5.9e-01	5.9e-01	1.4e-01	2.9	2.4e-01	2.3
stomach	1	4.7e-01	1	1.0	6.4e-01	1.5
uterus	1	2.4e-01	1	1.0	1.7e-01	2.0

As noted above, cluster AA161187 features 7 transcript(s), which were listed in Table 403 above. These transcript(s) encode for protein(s) which are variant(s) of protein Testisin precursor. A description of each variant protein according to the present invention is now provided.

Variant protein AA161187_P1 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T0.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide.

Variant protein AA161187_P1 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 408, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 408 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
1	M ->	No
16	A ->	No
226	N ->	No
253	I -> V	No
255	V -> I	No
264	R ->	No
264	R -> P	No
264	R -> Q	Yes

Variant protein AA161187_P1 is encoded by the following transcript(s): AA161187_T0, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T0 is shown in bold; this coding portion starts at position 107 and ends at position 1048. The transcript also has the following SNPs as listed in Table 409 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 409 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
66	T -> A	No
67	T -> G	No
105	C -> T	No
108	T ->	No
154	T ->	No
190	C -> G	No
469	A -> G	Yes
571	C -> T	Yes
782	A ->	No
859	T -> C	Yes
863	A -> G	No
869	G -> A	No
897	G ->	No
897	G -> A	Yes
897	G -> C	No
1000	A -> G	Yes
1068	G ->	No
1068	G -> A	No
1069	C -> A	No
1168	A -> G	Yes

Variant protein AA161187_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T7. An alignment is given to the known protein (Testisin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of

the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between AA161187_P6 and TEST_HUMAN:

1.An isolated chimeric polypeptide encoding for AA161187_P6, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR corresponding to amino acids 1 - 42 of AA161187_P6, and a second amino acid sequence being at least 90 % homologous to GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYS DLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPV TYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNH LFLKYSFRKDIFGDMVCAGNAQGGKDACFGDSGGPLACNKNGLWYQIGVVSWGVGC GRPNRPGVYTNISHHFEWIQKLMAQSGMSQPDPSWPLLFFPLLWALPLLGPV corresponding to amino acids 31 - 314 of TEST_HUMAN, which also corresponds to amino acids 43 - 326 of AA161187_P6, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of AA161187_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR of AA161187_P6.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein AA161187_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 410, (given according to their position(s) on the amino acid

sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 410 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
238	N ->	No
265	I -> V	No
267	V -> I	No
276	R ->	No
276	R -> P	No
276	R -> Q	Yes

The glycosylation sites of variant protein AA161187_P6, as compared to the known protein Testisin precursor, are described in Table 411 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 411 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
200	yes	212
167	yes	179
273	yes	285

Variant protein AA161187_P6 is encoded by the following transcript(s): AA161187_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T7 is shown in bold; this coding portion starts at position 1 and ends at position 979. The transcript also has the following SNPs as listed in Table 412 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last

column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 412 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
400	A -> G	Yes
502	C -> T	Yes
713	A ->	No
790	T -> C	Yes
794	A -> G	No
800	G -> A	No
828	G ->	No
828	G -> A	Yes
828	G -> C	No
931	A -> G	Yes
999	G ->	No
999	G -> A	No
1000	C -> A	No
1099	A -> G	Yes

Variant protein AA161187_P13 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T15. An alignment is given to the known protein (Testisin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between AA161187_P13 and TEST_HUMAN:

1.An isolated chimeric polypeptide encoding for AA161187_P13, comprising a first amino acid sequence being at least 90 % homologous to MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS LRLWDSHVCGVSLLSHRW ALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GSSGRHHKQLYVQPPLPQVQFPQGHLWRHG corresponding to amino acids 184 - 213 of AA161187_P13, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of AA161187_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GSSGRHHKQLYVQPPLPQVQFPQGHLWRHG in AA161187_P13.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein AA161187_P13 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 413, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 413 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
1	M ->	No
16	A ->	No

The glycosylation sites of variant protein AA161187_P13, as compared to the known protein Testisin precursor, are described in Table 414 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 414 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
200	no	
167	yes	167
273	no	

Variant protein AA161187_P13 is encoded by the following transcript(s): AA161187_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T15 is shown in bold; this coding portion starts at position 107 and ends at position 745. The transcript also has the following SNPs as listed in Table 415 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 415 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
66	T -> A	No

67	T -> G	No
105	C -> T	No
108	T ->	No
154	T ->	No
190	C -> G	No
469	A -> G	Yes
571	C -> T	Yes
791	T -> C	Yes
795	A -> G	No
801	G -> A	No
829	G ->	No
829	G -> A	Yes
829	G -> C	No
932	A -> G	Yes
1000	G ->	No
1000	G -> A	No
1001	C -> A	No
1100	A -> G	Yes

Variant protein AA161187_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T16. An alignment is given to the known protein (Testisin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between AA161187_P14 and TEST_HUMAN:

1.An isolated chimeric polypeptide encoding for AA161187_P14, comprising a first amino acid sequence being at least 90 % homologous to MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS LRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY

YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GCCLSPSHYRPHSTAISPHPPGSSGRHHKQLYVQPPLPQVQFPQGHLWRHGLCWQCPRR EGCLLRECPCHHSQPRKASCVPVPYLTLMPTPGGGDCCPTLQMQKRRLGCCQGEEEDV HPVYPAP corresponding to amino acids 184 - 307 of AA161187_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of AA161187_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GCCLSPSHYRPHSTAISPHPPGSSGRHHKQLYVQPPLPQVQFPQGHLWRHGLCWQCPRR EGCLLRECPCHHSQPRKASCVPVPYLTLMPTPGGGDCCPTLQMQKRRLGCCQGEEEDV HPVYPAP in AA161187_P14.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein AA161187_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 416, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 416 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
1	M ->	No
16	A ->	No
238	Q ->	No

The glycosylation sites of variant protein AA161187_P14, as compared to the known protein Testisin precursor, are described in Table 417 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 417 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
200	no	
167	yes	167
273	no	

Variant protein AA161187_P14 is encoded by the following transcript(s): AA161187_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T16 is shown in bold; this coding portion starts at position 107 and ends at position 1027. The transcript also has the following SNPs as listed in Table 418 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 418 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?

66	T -> A	No
67	T -> G	No
105	C -> T	No
108	T ->	No
154	T ->	No
190	C -> G	No
469	A -> G	Yes
571	C -> T	Yes
819	A ->	No
859	C -> T	Yes
1152	T -> C	Yes
1156	A -> G	No
1162	G -> A	No
1190	G ->	No
1190	G -> A	Yes
1190	G -> C	No
1293	A -> G	Yes
1361	G ->	No
1361	G -> A	No
1362	C -> A	No
1461	A -> G	Yes

Variant protein AA161187_P18 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T20. An alignment is given to the known protein (Testisin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between AA161187_P18 and TEST_HUMAN:

1.An isolated chimeric polypeptide encoding for AA161187_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR corresponding to amino acids 1 - 42 of AA161187_P18, a second amino acid sequence being at least 90 % homologous to GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFET corresponding to amino acids 31 - 86 of TEST_HUMAN, which also corresponds to amino acids 43 - 98 of AA161187_P18, a third amino acid sequence being at least 90 % homologous to DLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPV TYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNH LFLKYSFRKDIFGDMVCAGNAQGGKDACF corresponding to amino acids 89 - 235 of TEST_HUMAN, which also corresponds to amino acids 99 - 245 of AA161187_P18, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSVPATTPSPGKHPVSLCLI corresponding to amino acids 246 - 265 of AA161187_P18, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of AA161187_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIR of AA161187_P18.

3.An isolated chimeric polypeptide encoding for an edge portion of AA161187_P18, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise TD, having a structure as follows: a sequence starting from any of amino acid numbers 98-x to 98; and ending at any of amino acid numbers 99 + ((n-2) - x), in which x varies from 0 to n-2.

4.An isolated polypeptide encoding for a tail of AA161187_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSVPATTPSPGKHPVSLCLI in AA161187_P18.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein AA161187_P18 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 419, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P18 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 419 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
236	N ->	No
249	P -> L	Yes

The glycosylation sites of variant protein AA161187_P18, as compared to the known protein Testisin precursor, are described in Table 420 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 420 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
200	yes	210
167	yes	177
273	no	

Variant protein AA161187_P18 is encoded by the following transcript(s): AA161187_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T20 is shown in bold; this coding portion starts at position 1 and ends at position 796. The transcript also has the following SNPs as listed in Table 421 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P18 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 421 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
394	A -> G	Yes
496	C -> T	Yes
707	A ->	No
747	C -> T	Yes
1040	T -> C	Yes
1044	A -> G	No
1050	G -> A	No
1078	G ->	No
1078	G -> A	Yes
1078	G -> C	No
1181	A -> G	Yes
1249	G ->	No

1249	G -> A	No
1250	C -> A	No
1349	A -> G	Yes

Variant protein AA161187_P19 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AA161187_T21. An alignment is given to the known protein (Testisin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between AA161187_P19 and TEST_HUMAN:

1.An isolated chimeric polypeptide encoding for AA161187_P19, comprising a first amino acid sequence being at least 90 % homologous to MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGS LRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAY YTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTG WGYIKEDE corresponding to amino acids 1 - 183 of TEST_HUMAN, which also corresponds to amino acids 1 - 183 of AA161187_P19, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DKRTQ corresponding to amino acids 184 - 188 of AA161187_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of AA161187_P19, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DKRTQ in AA161187_P19.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein AA161187_P19 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 422, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 422 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
1	M ->	No
16	A ->	No

The glycosylation sites of variant protein AA161187_P19, as compared to the known protein Testisin precursor, are described in Table 423 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 423 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
200	no	
167	yes	167
273	no	

Variant protein AA161187_P19 is encoded by the following transcript(s): AA161187_T21, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AA161187_T21 is shown in bold; this coding portion starts at position 107

and ends at position 670. The transcript also has the following SNPs as listed in Table 424 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AA161187_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 424 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
66	T -> A	No
67	T -> G	No
105	C -> T	No
108	T ->	No
154	T ->	No
190	C -> G	No
469	A -> G	Yes
571	C -> T	Yes
719	G -> T	Yes

As noted above, cluster AA161187 features 20 segment(s), which were listed in Table 404 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster AA161187_node_0 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T15, AA161187_T16, AA161187_T21 and AA161187_T22. Table 425 below describes the starting and ending position of this segment on each transcript.

Table 425 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	1	170
AA161187_T15	1	170
AA161187_T16	1	170
AA161187_T21	1	170
AA161187_T22	1	170

Segment cluster AA161187_node_6 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T7 and AA161187_T20. Table 426 below describes the starting and ending position of this segment on each transcript.

Table 426 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T7	1	120
AA161187_T20	1	120

Segment cluster AA161187_node_14 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T20, AA161187_T21 and AA161187_T22. Table 427 below describes the starting and ending position of this segment on each transcript.

Table 427 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	446	656

AA161187_T7	377	587
AA161187_T15	446	656
AA161187_T16	446	656
AA161187_T20	371	581
AA161187_T21	446	656
AA161187_T22	446	656

Segment cluster AA161187_node_16 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T22. Table 428 below describes the starting and ending position of this segment on each transcript.

Table 428 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T22	657	953

Segment cluster AA161187_node_25 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T16 and AA161187_T20. Table 429 below describes the starting and ending position of this segment on each transcript.

Table 429 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T16	880	1104
AA161187_T20	768	992

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 430.

Table 430 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
AA161187_0_0_430	lung malignant tumors	LUN

Segment cluster AA161187_node_26 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16 and AA161187_T20. Table 431 below describes the starting and ending position of this segment on each transcript.

Table 431 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	812	1173
AA161187_T7	743	1104
AA161187_T15	744	1105
AA161187_T16	1105	1466
AA161187_T20	993	1354

Segment cluster AA161187_node_28 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T21. Table 432 below describes the starting and ending position of this segment on each transcript.

Table 432 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T21	657	1171

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster AA161187_node_4 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T15, AA161187_T16, AA161187_T21 and AA161187_T22. Table 433 below describes the starting and ending position of this segment on each transcript.

Table 433 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	171	197
AA161187_T15	171	197
AA161187_T16	171	197
AA161187_T21	171	197
AA161187_T22	171	197

Segment cluster AA161187_node_7 according to the present invention can be found in the following transcript(s): AA161187_T7 and AA161187_T20. Table 434 below describes the starting and ending position of this segment on each transcript.

Table 434 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T7	121	128

AA161187_T20	121	128

Segment cluster AA161187_node_8 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T20, AA161187_T21 and AA161187_T22. Table 435 below describes the starting and ending position of this segment on each transcript.

Table 435 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	198	256
AA161187_T7	129	187
AA161187_T15	198	256
AA161187_T16	198	256
AA161187_T20	129	187
AA161187_T21	198	256
AA161187_T22	198	256

Segment cluster AA161187_node_9 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T20, AA161187_T21 and AA161187_T22. Table 436 below describes the starting and ending position of this segment on each transcript.

Table 436 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	257	298

AA161187_T7	188	229
AA161187_T15	257	298
AA161187_T16	257	298
AA161187_T20	188	229
AA161187_T21	257	298
AA161187_T22	257	298

Segment cluster AA161187_node_10 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T20, AA161187_T21 and AA161187_T22. Table 437 below describes the starting and ending position of this segment on each transcript.

Table 437 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	299	363
AA161187_T7	230	294
AA161187_T15	299	363
AA161187_T16	299	363
AA161187_T20	230	294
AA161187_T21	299	363
AA161187_T22	299	363

Segment cluster AA161187_node_12 according to the present invention can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T21 and AA161187_T22. Table 438 below describes the starting and ending position of this segment on each transcript.

Table 438 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	364	369
AA161187_T7	295	300
AA161187_T15	364	369
AA161187_T16	364	369
AA161187_T21	364	369
AA161187_T22	364	369

Segment cluster AA161187_node_13 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16, AA161187_T20, AA161187_T21 and AA161187_T22. Table 439 below describes the starting and ending position of this segment on each transcript.

Table 439 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	370	445
AA161187_T7	301	376
AA161187_T15	370	445
AA161187_T16	370	445
AA161187_T20	295	370
AA161187_T21	370	445
AA161187_T22	370	445

Segment cluster AA161187_node_19 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): AA161187_T16. Table 440 below describes the starting and ending position of this segment on each transcript.

Table 440 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T16	657	693

Segment cluster AA161187_node_20 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T16 and AA161187_T20. Table 441 below describes the starting and ending position of this segment on each transcript.

Table 441 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	657	682
AA161187_T7	588	613
AA161187_T16	694	719
AA161187_T20	582	607

Segment cluster AA161187_node_21 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16 and AA161187_T20. Table 442 below describes the starting and ending position of this segment on each transcript.

Table 442 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	683	741
AA161187_T7	614	672
AA161187_T15	657	715
AA161187_T16	720	778
AA161187_T20	608	666

Segment cluster AA161187_node_22 according to the present invention is supported by 34 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T15, AA161187_T16 and AA161187_T20. Table 443 below describes the starting and ending position of this segment on each transcript.

Table 443 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	742	769
AA161187_T7	673	700
AA161187_T15	716	743
AA161187_T16	779	806
AA161187_T20	667	694

Segment cluster AA161187_node_23 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T0, AA161187_T7, AA161187_T16 and AA161187_T20. Table 444 below describes the starting and ending position of this segment on each transcript.

Table 444 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T0	770	811
AA161187_T7	701	742
AA161187_T16	807	848
AA161187_T20	695	736

Segment cluster AA161187_node_24 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AA161187_T16 and AA161187_T20. Table 445 below describes the starting and ending position of this segment on each transcript.

Table 445 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AA161187_T16	849	879
AA161187_T20	737	767

Variant protein alignment to the previously known protein:

Sequence name: TEST_HUMAN

Sequence documentation:

Alignment of: AA161187_P6 x TEST_HUMAN ..

Alignment segment 1/1:

Quality: 2894.00

Escore: 0

Matching length: 284 Total length: 284

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      43 GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTA 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTA 80
                .         .         .         .         .
      93 AHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRY 142
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 AHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRY 130
                .         .         .         .         .
     143 LGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIK 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 LGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIK 180
                .         .         .         .         .
     193 EDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGG 242
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 EDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGG 230
```

```
                .         .         .         .         .
      243 KDACFGDSGGPLACNKNGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEW 292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      231 KDACFGDSGGPLACNKNGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEW 280
                    .         .         .
      293 IQKLMAQSGMSQPDPSWPLLFFPLLWALPLLGPV                  326
          |||||||||||||||||||||||||||||||||
      281 IQKLMAQSGMSQPDPSWPLLFFPLLWALPLLGPV                  314
```

Sequence name: TEST_HUMAN

Sequence documentation:

Alignment of: AA161187_P13 x TEST_HUMAN ..

Alignment segment 1/1:

Quality: 1829.00

Escore: 0

Matching length: 183 Total length: 183

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                 .         .         .         .         .
       1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
                 .         .         .         .         .
      51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
                 .         .         .         .         .
     101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
                 .         .         .
     151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                   183
         |||||||||||||||||||||||||||||||||
     151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                   183
```

Sequence name: TEST_HUMAN

Sequence documentation:

Alignment of: AA161187_P14 x TEST_HUMAN ..

Alignment segment 1/1:

Quality: 1829.00

Escore: 0

Matching length: 183 Total length: 183

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
                .         .         .         .         .
       51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
                .         .         .         .         .
      101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
                .         .         .
      151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                  183
          |||||||||||||||||||||||||||||||||
      151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                  183
```

Sequence name: TEST_HUMAN

Sequence documentation:

Alignment of: AA161187_P18 x TEST_HUMAN ..

Alignment segment 1/1:

Quality: 1957.00
Escore: 0
Matching length: 203 Total length: 205
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 99.02 Total Percent Identity: 99.02
Gaps: 1

Alignment:

```
                .         .         .         .         .
      43 GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTA 92
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 GPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWALTA 80
                .         .         .         .         .
      93 AHCFET..DLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRY 140
         ||||||  ||||||||||||||||||||||||||||||||||||||||||||
      81 AHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRY 130
                .         .         .         .         .
```

545

```
141 LGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIK 190
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
131 LGNSPYDIALVKLSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIK 180
              .         .         .         .         .
191 EDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGG 240
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
181 EDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGG 230

241 KDACF                                               245
    |||||
231 KDACF                                               235
```

Sequence name: TEST_HUMAN

Sequence documentation:

Alignment of: AA161187_P19 x TEST_HUMAN ..

Alignment segment 1/1:

Quality: 1829.00

Escore: 0

Matching length: 183 Total length: 183

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

546

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAEL 50
                .         .         .         .         .
     51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GRWPWQGSLRLWDSHVCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQF 100
                .         .         .         .         .
    101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 GQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYT 150
                .         .         .
    151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                183
        |||||||||||||||||||||||||||||||||
    151 KHIQPICLQASTFEFENRTDCWVTGWGYIKEDE                183
```

Expression of Homo sapiens protease, serine, 21 (testisin) (PRSS21) AA161187 transcripts which are detectable by amplicon as depicted in sequence name AA161187 seg25 in normal and cancerous lung tissues

Expression of Homo sapiens protease, serine, 21 (testisin) (PRSS21) transcripts detectable by or according to seg25, AA161187 seg25 amplicon (SEQ ID NO:1654) and primers AA161187 seg17F2 (SEQ ID NO:1652) and AA161187 seg17R2 (SEQ ID NO:1653) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1

(GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 64 is a histogram showing over expression of the above-indicated Homo sapiens protease, serine, 21 (testisin) (PRSS21) transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 64, the expression of Homo sapiens protease, serine, 21 (testisin) (PRSS21) transcripts detectable by the above amplicon(s) was higher in a few cancer samples than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 6 fold was found in 1 out of 15 adenocarcinoma samples, 3 out of 16 squamous cell carcinoma samples, 1 out of 4 large cell carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: AA161187 seg17F2 forward primer; and AA161187 seg17R2 reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: AA161187 seg25.

Primers:

Forward primer AA161187 seg17F2 (SEQ ID NO:1652):

CCCTGTGCCTTATTTGACCCT

Reverse primer AA161187 seg17R2 (SEQ ID NO:1653) :

GCTGGGTAGACTGGGTGCA

Amplicon AA161187 seg25 (SEQ ID NO:1654):

CCTGTGCCTTATTTGACCCTCATGCCAACCCCGGGAGGTGGAGACTGTTGCCCCACT
CTGCAGATGCAGAAACGGAGGCTTGGCTGCTGCCAGGGGGAGGA

DESCRIPTION FOR CLUSTER R66178

Cluster R66178 features 3 transcript(s) and 16 segment(s) of interest, the names for which are given in Tables 446 and 447, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 448.

Table 446 - Transcripts of interest

Transcript Name	Sequence ID No.
R66178_T2	48
R66178_T3	49
R66178_T7	50

Table 447 - Segments of interest

Segment Name	Sequence ID No.
R66178_node_0	502
R66178_node_6	503
R66178_node_8	504
R66178_node_15	505
R66178_node_24	506
R66178_node_26	507
R66178_node_27	508
R66178_node_4	509
R66178_node_5	510
R66178_node_9	511
R66178_node_11	512
R66178_node_16	513
R66178_node_18	514
R66178_node_19	515
R66178_node_20	516
R66178_node_21	517

Table 448 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
R66178_P3	1324	R66178_T2
R66178_P4	1325	R66178_T3
R66178_P8	1326	R66178_T7

These sequences are variants of the known protein Poliovirus receptor related protein 1 precursor (SwissProt accession identifier PVR1_HUMAN; known also according to the synonyms Herpes virus entry mediator C; HveC; Nectin 1; Herpesvirus Ig-like receptor; HIgR; CD111 antigen), SEQ ID NO: 1432, referred to herein as the previously known protein.

Protein Poliovirus receptor related protein 1 precursor is known or believed to have the following function(s): probably involved in cell adhesion; receptor for alphaherpesvirus (HSV-

1, HSV-2 and Pseudorabies virus) entry into cells. The sequence for protein Poliovirus receptor related protein 1 precursor is given at the end of the application, as "Poliovirus receptor related protein 1 precursor amino acid sequence". Protein Poliovirus receptor related protein 1 precursor localization is believed to be Type I membrane protein (isoforms alpha and delta). Secreted (isoform gamma).

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: immune response; cell-cell adhesion, which are annotation(s) related to Biological Process; cell adhesion receptor; protein binding; coreceptor, which are annotation(s) related to Molecular Function; and adherens junction; integral membrane protein, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster R66178 features 3 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Poliovirus receptor related protein 1 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein R66178_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R66178_T2. An alignment is given to the known protein (Poliovirus receptor related protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R66178_P3 and PVR1_HUMAN:

1.An isolated chimeric polypeptide encoding for R66178_P3, comprising a first amino acid sequence being at least 90 % homologous to

MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP
LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLEL
EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS
ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM

DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT corresponding to amino acids 1 - 334 of PVR1_HUMAN, which also corresponds to amino acids 1 - 334 of R66178_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GEGHSLPISPGVLQTQNCGP corresponding to amino acids 335 - 354 of R66178_P3, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R66178_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GEGHSLPISPGVLQTQNCGP in R66178_P3.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R66178_P3 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 449, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 449 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
77	N -> S	No

The glycosylation sites of variant protein R66178_P3, as compared to the known protein Poliovirus receptor related protein 1 precursor, are described in Table 450 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 450 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
72	yes	72
297	yes	297
202	yes	202
307	yes	307
332	yes	332
139	yes	139
36	yes	36
286	yes	286

Variant protein R66178_P3 is encoded by the following transcript(s): R66178_T2, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R66178_T2 is shown in bold; this coding portion starts at position 634 and ends at position 1695. The transcript also has the following SNPs as listed in Table 451 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 451 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
474	-> T	No

476	-> C	No
632	-> T	No
633	G -> T	No
863	A -> G	No
897	C -> T	Yes
2178	A -> G	No
2465	G -> A	Yes
2687	G -> A	Yes

Variant protein R66178_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R66178_T3. An alignment is given to the known protein (Poliovirus receptor related protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R66178_P4 and PVR1_HUMAN:

1.An isolated chimeric polypeptide encoding for R66178_P4, comprising a first amino acid sequence being at least 90 % homologous to MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLEL EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT corresponding to amino acids 1 - 334 of PVR1_HUMAN, which also corresponds to amino acids 1 - 334 of R66178_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence AFCQLIYPGKGRTRARMF corresponding to amino acids 335 - 352 of R66178_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R66178_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence AFCQLIYPGKGRTRARMF in R66178_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R66178_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 452, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 452 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
77	N -> S	No

The glycosylation sites of variant protein R66178_P4, as compared to the known protein Poliovirus receptor related protein 1 precursor, are described in Table 453 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 453 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?

72	yes	72
297	yes	297
202	yes	202
307	yes	307
332	yes	332
139	yes	139
36	yes	36
286	yes	286

Variant protein R66178_P4 is encoded by the following transcript(s): R66178_T3, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R66178_T3 is shown in bold; this coding portion starts at position 634 and ends at position 1689. The transcript also has the following SNPs as listed in Table 454 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 454 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
474	-> T	No
476	-> C	No
632	-> T	No
633	G -> T	No
863	A -> G	No
897	C -> T	Yes
1762	C ->	Yes

Variant protein R66178_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R66178_T7. An alignment is given to the known protein (Poliovirus receptor related protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R66178_P8 and PVR1_HUMAN:

1.An isolated chimeric polypeptide encoding for R66178_P8, comprising a first amino acid sequence being at least 90 % homologous to MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANP LPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLEL EDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQAVLRAKKGQDDKVLVATCTS ANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYHM DRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLP KGVEAQNRTLFFKGPINYSLAGTYICEATNPIGTRSGQVE corresponding to amino acids 1 - 330 of PVR1_HUMAN, which also corresponds to amino acids 1 - 330 of R66178_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence NSPTPRLLPNMGGAPGRCPRPSLGAWRGASCWC corresponding to amino acids 331 - 363 of R66178_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R66178_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence NSPTPRLLPNMGGAPGRCPRPSLGAWRGASCWC in R66178_P8.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide

prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R66178_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 455, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 455 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
77	N -> S	No

The glycosylation sites of variant protein R66178_P8, as compared to the known protein Poliovirus receptor related protein 1 precursor, are described in Table 456 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 456 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
72	yes	72
297	yes	297
202	yes	202
307	yes	307
332	no	
139	yes	139
36	yes	36
286	yes	286

Variant protein R66178_P8 is encoded by the following transcript(s): R66178_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R66178_T7 is shown in bold; this coding portion starts at position 634 and ends at position 1722. The transcript also has the following SNPs as listed in Table 457 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R66178_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 457 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
474	-> T	No
476	-> C	No
632	-> T	No
633	G -> T	No
863	A -> G	No
897	C -> T	Yes
2210	A -> C	No
2211	A -> C	No

As noted above, cluster R66178 features 16 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R66178_node_0 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 458 below describes the starting and ending position of this segment on each transcript.

Table 458 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1	712
R66178_T3	1	712
R66178_T7	1	712

Segment cluster R66178_node_6 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 459 below describes the starting and ending position of this segment on each transcript.

Table 459 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	762	1063
R66178_T3	762	1063
R66178_T7	762	1063

Segment cluster R66178_node_8 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 460 below describes the starting and ending position of this segment on each transcript.

Table 460 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1064	1269
R66178_T3	1064	1269
R66178_T7	1064	1269

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 461.

Table 461 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
R66178_0_7_0	lung malignant tumors	LUN

Segment cluster R66178_node_15 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 462 below describes the starting and ending position of this segment on each transcript.

Table 462 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1485	1623
R66178_T3	1485	1623
R66178_T7	1485	1623

Segment cluster R66178_node_24 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2. Table 463 below describes the starting and ending position of this segment on each transcript.

Table 463 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

R66178_T2	1637	3110

Segment cluster R66178_node_26 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T7. Table 464 below describes the starting and ending position of this segment on each transcript.

Table 464 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T7	1624	2087

Segment cluster R66178_node_27 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T7. Table 465 below describes the starting and ending position of this segment on each transcript.

Table 465 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T7	2088	2364

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R66178_node_4 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 466 below describes the starting and ending position of this segment on each transcript.

Table 466 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	713	749
R66178_T3	713	749
R66178_T7	713	749

Segment cluster R66178_node_5 according to the present invention can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 467 below describes the starting and ending position of this segment on each transcript.

Table 467 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	750	761
R66178_T3	750	761
R66178_T7	750	761

Segment cluster R66178_node_9 according to the present invention is supported by 44 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 468 below describes the starting and ending position of this segment on each transcript.

Table 468 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1270	1366
R66178_T3	1270	1366
R66178_T7	1270	1366

Segment cluster R66178_node_11 according to the present invention is supported by 44 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T2, R66178_T3 and R66178_T7. Table 469 below describes the starting and ending position of this segment on each transcript.

Table 469 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1367	1484
R66178_T3	1367	1484
R66178_T7	1367	1484

Segment cluster R66178_node_16 according to the present invention can be found in the following transcript(s): R66178_T2 and R66178_T3. Table 470 below describes the starting and ending position of this segment on each transcript.

Table 470 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T2	1624	1636
R66178_T3	1624	1636

Segment cluster R66178_node_18 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T3. Table 471 below describes the starting and ending position of this segment on each transcript.

Table 471 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T3	1637	1743

Segment cluster R66178_node_19 according to the present invention can be found in the following transcript(s): R66178_T3. Table 472 below describes the starting and ending position of this segment on each transcript.

Table 472 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T3	1744	1763

Segment cluster R66178_node_20 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T3. Table 473 below describes the starting and ending position of this segment on each transcript.

Table 473 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T3	1764	1791

Segment cluster R66178_node_21 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R66178_T3. Table 474 below describes the starting and ending position of this segment on each transcript.

Table 474 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R66178_T3	1792	1903

Variant protein alignment to the previously known protein:

Sequence name: PVR1_HUMAN

Sequence documentation:

Alignment of: R66178_P3 x PVR1_HUMAN ..

Alignment segment 1/1:

Quality: 3286.00

Escore: 0

Matching length: 334 Total length: 334

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50

```
                                                                    
                                                                    
  1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50
                .         .         .         .         .
 51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
                                                                    
 51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
                .         .         .         .         .
101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
                                                                    
101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
                .         .         .         .         .
151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
                                                                    
151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
                .         .         .         .         .
201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
                                                                    
201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
                .         .         .         .         .
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
                                                                    
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
                .         .         .
301 FFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT                  334
                                                                    
301 FFKGPINYSLAGTYICEATNPIGTRSGQVEVNIT                  334
```

Sequence name: PVR1_HUMAN

Sequence documentation:

Alignment of: R66178_P4 x PVR1_HUMAN ..

Alignment segment 1/1:

Quality: 3294.00
Escore: 0
Matching length: 336 Total length: 336
Matching Percent Similarity: 99.70 Matching Percent Identity: 99.70
Total Percent Similarity: 99.70 Total Percent Identity: 99.70
Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50
                .         .         .         .         .
     51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
                .         .         .         .         .
    101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
                .         .         .         .         .
151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
                .         .         .         .         .
201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
                .         .         .         .         .
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
                .         .         .
301 FFKGPINYSLAGTYICEATNPIGTRSGQVEVNITAF                336
    |||||||||||||||||||||||||||||||| |
301 FFKGPINYSLAGTYICEATNPIGTRSGQVEVNITEF                336
```

Sequence name: PVR1_HUMAN

Sequence documentation:

Alignment of: R66178_P8 x PVR1_HUMAN ..

Alignment segment 1/1:

Quality: 3250.00

Escore: 0

Matching length: 330 Total length: 330

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLH 50
                .         .         .         .         .
     51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CSFANPLPSVKITQVTWQKSTNGSKQNVAIYNPSMGVSVLAPYRERVEFL 100
                .         .         .         .         .
    101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 RPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWI 150
                .         .         .         .         .
    151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 EGTQAVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRN 200
                .         .         .         .         .
    201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 PNGTVTVISRYRLVPSREAHQQSLACIVNYHMDRFKESLTLNVQYEPEVT 250
```

570

```
                .         .         .         .         .
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 IEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTL 300
                .         .         .
301 FFKGPINYSLAGTYICEATNPIGTRSGQVE                      330
    |||||||||||||||||||||||||||||
301 FFKGPINYSLAGTYICEATNPIGTRSGQVE                      330
```

DESCRIPTION FOR CLUSTER HUMPHOSLIP

Cluster HUMPHOSLIP features 7 transcript(s) and 53 segment(s) of interest, the names for which are given in Tables 475 and 476, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 477.

Table 475 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMPHOSLIP_PEA_2_T6	51
HUMPHOSLIP_PEA_2_T7	52
HUMPHOSLIP_PEA_2_T14	53
HUMPHOSLIP_PEA_2_T16	54
HUMPHOSLIP_PEA_2_T17	55
HUMPHOSLIP_PEA_2_T18	56
HUMPHOSLIP_PEA_2_T19	57

Table 476 - Segments of interest

Segment Name	Sequence ID No.
HUMPHOSLIP_PEA_2_node_0	518
HUMPHOSLIP_PEA_2_node_19	519
HUMPHOSLIP_PEA_2_node_34	520

HUMPHOSLIP_PEA_2_node_68	521
HUMPHOSLIP_PEA_2_node_70	522
HUMPHOSLIP_PEA_2_node_75	523
HUMPHOSLIP_PEA_2_node_2	524
HUMPHOSLIP_PEA_2_node_3	525
HUMPHOSLIP_PEA_2_node_4	526
HUMPHOSLIP_PEA_2_node_6	527
HUMPHOSLIP_PEA_2_node_7	528
HUMPHOSLIP_PEA_2_node_8	529
HUMPHOSLIP_PEA_2_node_9	530
HUMPHOSLIP_PEA_2_node_14	531
HUMPHOSLIP_PEA_2_node_15	532
HUMPHOSLIP_PEA_2_node_16	533
HUMPHOSLIP_PEA_2_node_17	534
HUMPHOSLIP_PEA_2_node_23	535
HUMPHOSLIP_PEA_2_node_24	536
HUMPHOSLIP_PEA_2_node_25	537
HUMPHOSLIP_PEA_2_node_26	538
HUMPHOSLIP_PEA_2_node_29	539
HUMPHOSLIP_PEA_2_node_30	540
HUMPHOSLIP_PEA_2_node_33	541
HUMPHOSLIP_PEA_2_node_36	542
HUMPHOSLIP_PEA_2_node_37	543
HUMPHOSLIP_PEA_2_node_39	544
HUMPHOSLIP_PEA_2_node_40	545
HUMPHOSLIP_PEA_2_node_41	546
HUMPHOSLIP_PEA_2_node_42	547
HUMPHOSLIP_PEA_2_node_44	548
HUMPHOSLIP_PEA_2_node_45	549
HUMPHOSLIP_PEA_2_node_47	550

HUMPHOSLIP_PEA_2_node_51	551
HUMPHOSLIP_PEA_2_node_52	552
HUMPHOSLIP_PEA_2_node_53	553
HUMPHOSLIP_PEA_2_node_54	554
HUMPHOSLIP_PEA_2_node_55	555
HUMPHOSLIP_PEA_2_node_58	556
HUMPHOSLIP_PEA_2_node_59	557
HUMPHOSLIP_PEA_2_node_60	558
HUMPHOSLIP_PEA_2_node_61	559
HUMPHOSLIP_PEA_2_node_62	560
HUMPHOSLIP_PEA_2_node_63	562
HUMPHOSLIP_PEA_2_node_64	562
HUMPHOSLIP_PEA_2_node_65	563
HUMPHOSLIP_PEA_2_node_66	564
HUMPHOSLIP_PEA_2_node_67	565
HUMPHOSLIP_PEA_2_node_69	566
HUMPHOSLIP_PEA_2_node_71	567
HUMPHOSLIP_PEA_2_node_72	568
HUMPHOSLIP_PEA_2_node_73	569
HUMPHOSLIP_PEA_2_node_74	570

Table 477 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HUMPHOSLIP_PEA_2_P10	1327	HUMPHOSLIP_PEA_2_T17
HUMPHOSLIP_PEA_2_P12	1328	HUMPHOSLIP_PEA_2_T19
HUMPHOSLIP_PEA_2_P30	1329	HUMPHOSLIP_PEA_2_T6
HUMPHOSLIP_PEA_2_P31	1330	HUMPHOSLIP_PEA_2_T7
HUMPHOSLIP_PEA_2_P33	1331	HUMPHOSLIP_PEA_2_T14
HUMPHOSLIP_PEA_2_P34	1332	HUMPHOSLIP_PEA_2_T16

HUMPHOSLIP_PEA_2_P35	1333	HUMPHOSLIP_PEA_2_T18

These sequences are variants of the known protein Phospholipid transfer protein precursor (SwissProt accession identifier PLTP_HUMAN; known also according to the synonyms Lipid transfer protein II), SEQ ID NO: 1433, referred to herein as the previously known protein.

Protein Phospholipid transfer protein precursor is known or believed to have the following function(s): Converts HDL into larger and smaller particles. May play a key role in extracellular phospholipid transport and modulation of HDL particles. The sequence for protein Phospholipid transfer protein precursor is given at the end of the application, as "Phospholipid transfer protein precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 478.

Table 478 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
282	R -> Q. /FTId=VAR_017020.
372	R -> H. /FTId=VAR_017021.
380	R -> W (in dbSNP:6065903). /FTId=VAR_017022.
444	F -> L (in dbSNP:1804161). /FTId=VAR_012073.
487	T -> K (in dbSNP:1056929). /FTId=VAR_012074.
18	E -> V

Protein Phospholipid transfer protein precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: lipid metabolism; lipid transport, which are annotation(s) related to Biological Process; lipid binding, which are annotation(s) related to Molecular Function; and extracellular, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

For this cluster, at least one oligonucleotide was found to demonstrate overexpression of the cluster, although not of at least one transcript/segment as listed below. Microarray (chip) data is also available for this cluster as follows. Various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer, as previously described. The following oligonucleotides were found to hit this cluster but not other segments/transcripts below, shown in Table 479, with regard to lung cancer.

Table 479 - Oligonucleotides related to this cluster

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMPHOSLIP_0_0_18458	lung malignant tumors	LUN

As noted above, cluster HUMPHOSLIP features 7 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Phospholipid transfer protein precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HUMPHOSLIP_PEA_2_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T17. An alignment is given to the known protein (Phospholipid transfer protein precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P10 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P10, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISE corresponding to amino acids 1 - 67 of PLTP_HUMAN, which also corresponds to amino acids 1 - 67 of HUMPHOSLIP_PEA_2_P10, and a second amino acid sequence being at least 90 % homologous to KVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMK DPVASTSNLDMDFRGAFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMES

YFRAGALQLLLVGDKVPHDLDMLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKP SGTTISVTASVTIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSN HSALESLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVTNHAGFLTI GADLHFAKGLREVIEKNRPADVRASTAPTPSTAAV corresponding to amino acids 163 - 493 of PLTP_HUMAN, which also corresponds to amino acids 68 - 398 of HUMPHOSLIP_PEA_2_P10, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HUMPHOSLIP_PEA_2_P10, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EK, having a structure as follows: a sequence starting from any of amino acid numbers 67-x to 67; and ending at any of amino acid numbers 68+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P10 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 480, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 480 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes
113	S -> F	Yes
118	V ->	No
140	R ->	No
140	R -> P	No
150	N ->	No
160	P ->	No
201	P ->	No
274	M ->	No
285	R -> W	Yes
292	Q ->	No
315	L -> *	No
330	M -> I	Yes
349	F -> L	Yes
392	T -> K	Yes

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P10, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 481 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 481 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	no	
143	no	

64	yes	64
245	yes	150
398	yes	303
117	no	

Variant protein HUMPHOSLIP_PEA_2_P10 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T17, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T17 is shown in bold; this coding portion starts at position 276 and ends at position 1469. The transcript also has the following SNPs as listed in Table 482 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 482 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
431	G -> A	Yes
551	C -> T	Yes
613	C -> T	Yes
628	T ->	No
694	G ->	No
694	G -> C	No
723	A ->	No
753	C ->	No
876	C ->	No

1037	C -> T	Yes
1097	G ->	No
1128	C -> T	Yes
1149	C ->	No
1219	T -> A	No
1230	C -> T	Yes
1265	G -> C	Yes
1322	T -> A	Yes
1450	C -> A	Yes
1469	C -> T	No
1549	C -> T	Yes
1565	A -> G	No
1565	A -> T	No
1630	A -> G	Yes
1654	T -> A	No
1731	G -> T	Yes
1864	G -> A	Yes
1893	G -> T	Yes
2073	G -> A	Yes
2269	C -> T	Yes
2325	G -> T	Yes
2465	C -> T	Yes
2566	C -> T	Yes
2881	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P12 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T19. An alignment is given to the known protein (Phospholipid transfer protein precursor) at the end of the application. One or more alignments to one or more

previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P12 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P12, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRG AFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDK VPHDLDMLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASVTIALVP PDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHSALESLALIPLQAPLK TMLQIGVMPMLN corresponding to amino acids 1 - 427 of PLTP_HUMAN, which also corresponds to amino acids 1 - 427 of HUMPHOSLIP_PEA_2_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKAGV corresponding to amino acids 428 - 432 of HUMPHOSLIP_PEA_2_P12, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKAGV in HUMPHOSLIP_PEA_2_P12.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P12 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 483, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 483 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes
81	D -> H	Yes
124	S -> Y	Yes
160	T ->	No
160	T -> N	No
208	S -> F	Yes
213	V ->	No
235	R -> P	No
235	R ->	No
245	N ->	No
255	P ->	No
296	P ->	No
369	M ->	No
380	R -> W	Yes
387	Q ->	No
410	L -> *	No
425	M -> I	Yes

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P12, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 484 (given

according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 484 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	yes	94
143	yes	143
64	yes	64
245	yes	245
398	yes	398
117	yes	117

Variant protein HUMPHOSLIP_PEA_2_P12 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T19, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T19 is shown in bold; this coding portion starts at position 276 and ends at position 1571. The transcript also has the following SNPs as listed in Table 485 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 485 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes

431	G -> A	Yes
516	G -> C	Yes
644	G -> A	Yes
646	C -> A	Yes
754	C ->	No
754	C -> A	No
836	C -> T	Yes
898	C -> T	Yes
913	T ->	No
979	G ->	No
979	G -> C	No
1008	A ->	No
1038	C ->	No
1161	C ->	No
1322	C -> T	Yes
1382	G ->	No
1413	C -> T	Yes
1434	C ->	No
1504	T -> A	No
1515	C -> T	Yes
1550	G -> C	Yes
1690	T -> A	Yes
1818	C -> A	Yes
1837	C -> T	No
1917	C -> T	Yes
1933	A -> G	No
1933	A -> T	No
1998	A -> G	Yes
2022	T -> A	No
2099	G -> T	Yes

2232	G -> A	Yes
2261	G -> T	Yes
2441	G -> A	Yes
2637	C -> T	Yes
2693	G -> T	Yes
2833	C -> T	Yes
2934	C -> T	Yes
3249	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P30 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T6. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P30 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 486, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P30 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 486 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes

37	R -> Q	Yes

Variant protein HUMPHOSLIP_PEA_2_P30 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T6 is shown in bold; this coding portion starts at position 276 and ends at position 431. The transcript also has the following SNPs as listed in Table 487 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P30 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 487- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
385	G -> A	Yes
470	G -> C	Yes
598	G -> A	Yes
600	C -> A	Yes
708	C ->	No
708	C -> A	No
790	C -> T	Yes
852	C -> T	Yes
867	T ->	No
933	G ->	No
933	G -> C	No
962	A ->	No
992	C ->	No

1115	C ->	No
1276	C -> T	Yes
1336	G ->	No
1367	C -> T	Yes
1388	C ->	No
1458	T -> A	No
1469	C -> T	Yes
1504	G -> C	Yes
1561	T -> A	Yes
1689	C -> A	Yes
1708	C -> T	No
1788	C -> T	Yes
1804	A -> G	No
1804	A -> T	No
1869	A -> G	Yes
1893	T -> A	No
1970	G -> T	Yes
2103	G -> A	Yes
2132	G -> T	Yes
2312	G -> A	Yes
2508	C -> T	Yes
2564	G -> T	Yes
2704	C -> T	Yes
2805	C -> T	Yes
3120	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P31 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T7. An alignment is given to the known protein (Phospholipid transfer

protein precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P31 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P31, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISE corresponding to amino acids 1 - 67 of PLTP_HUMAN, which also corresponds to amino acids 1 - 67 of HUMPHOSLIP_PEA_2_P31, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PGLERGADKFPVVGGSSLFLALDLTLRPPVG corresponding to amino acids 68 - 98 of HUMPHOSLIP_PEA_2_P31, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P31, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PGLERGADKFPVVGGSSLFLALDLTLRPPVG in HUMPHOSLIP_PEA_2_P31.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P31 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 488, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein

HUMPHOSLIP_PEA_2_P31 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 488 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P31, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 489(given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 489 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	no	
143	no	
64	yes	64
245	no	
398	no	
117	no	

Variant protein HUMPHOSLIP_PEA_2_P31 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T7 is shown in bold; this coding portion starts at position 276 and ends at position 569. The transcript also has the following SNPs as listed in Table 490 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not;

the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P31 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 490 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
431	G -> A	Yes
608	G -> C	Yes
736	G -> A	Yes
738	C -> A	Yes
846	C ->	No
846	C -> A	No
928	C -> T	Yes
990	C -> T	Yes
1005	T ->	No
1071	G ->	No
1071	G -> C	No
1100	A ->	No
1130	C ->	No
1253	C ->	No
1414	C -> T	Yes
1474	G ->	No
1505	C -> T	Yes
1526	C ->	No
1596	T -> A	No
1607	C -> T	Yes

1642	G -> C	Yes
1699	T -> A	Yes
1827	C -> A	Yes
1846	C -> T	No
1926	C -> T	Yes
1942	A -> G	No
1942	A -> T	No
2007	A -> G	Yes
2031	T -> A	No
2108	G -> T	Yes
2241	G -> A	Yes
2270	G -> T	Yes
2450	G -> A	Yes
2646	C -> T	Yes
2702	G -> T	Yes
2842	C -> T	Yes
2943	C -> T	Yes
3258	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P33 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T14. An alignment is given to the known protein (Phospholipid transfer protein precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P33 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P33, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH

FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQ corresponding to amino acids 1 - 183 of PLTP_HUMAN, which also corresponds to amino acids 1 - 183 of HUMPHOSLIP_PEA_2_P33, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VWAATGRRVARVGMLSL corresponding to amino acids 184 - 200 of HUMPHOSLIP_PEA_2_P33, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P33, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VWAATGRRVARVGMLSL in HUMPHOSLIP_PEA_2_P33.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P33 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 491, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P33 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 491 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes

18	E -> V	Yes
81	D -> H	Yes
124	S -> Y	Yes
160	T ->	No
160	T -> N	No

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P33, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 492 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 492 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	yes	94
143	yes	143
64	yes	64
245	no	
398	no	
117	yes	117

Variant protein HUMPHOSLIP_PEA_2_P33 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T14, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T14 is shown in bold; this coding portion starts at position 276 and ends at position 875. The transcript also has the following SNPs as listed in Table 493 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P33 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 493 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
431	G -> A	Yes
516	G -> C	Yes
644	G -> A	Yes
646	C -> A	Yes
754	C ->	No
754	C -> A	No
921	C -> T	Yes
983	C -> T	Yes
998	T ->	No
1064	G ->	No
1064	G -> C	No
1093	A ->	No
1123	C ->	No
1246	C ->	No
1407	C -> T	Yes
1467	G ->	No
1498	C -> T	Yes
1519	C ->	No
1589	T -> A	No
1600	C -> T	Yes
1635	G -> C	Yes
1692	T -> A	Yes
1820	C -> A	Yes

1839	C -> T	No
1919	C -> T	Yes
1935	A -> G	No
1935	A -> T	No
2000	A -> G	Yes
2024	T -> A	No
2101	G -> T	Yes
2234	G -> A	Yes
2263	G -> T	Yes
2443	G -> A	Yes
2639	C -> T	Yes
2695	G -> T	Yes
2835	C -> T	Yes
2936	C -> T	Yes
3251	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P34 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T16. An alignment is given to the known protein (Phospholipid transfer protein precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P34 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P34, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINAS AEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMHAAFGGTFKKVYDFLSTFITSGMRF LLNQQICPVLYHAGTVLLNSLLDTVPV corresponding to amino acids 1 - 205 of

PLTP_HUMAN, which also corresponds to amino acids 1 - 205 of HUMPHOSLIP_PEA_2_P34, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LWTSLLALTIPS corresponding to amino acids 206 - 217 of HUMPHOSLIP_PEA_2_P34, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P34, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LWTSLLALTIPS in HUMPHOSLIP_PEA_2_P34.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P34 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 494, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P34 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 494 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes
81	D -> H	Yes
124	S -> Y	Yes

160	T ->	No
160	T -> N	No
211	L ->	No

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P34, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 495 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 495 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	yes	94
143	yes	143
64	yes	64
245	no	
398	no	
117	yes	117

Variant protein HUMPHOSLIP_PEA_2_P34 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T16 is shown in bold; this coding portion starts at position 276 and ends at position 926. The transcript also has the following SNPs as listed in Table 496 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P34 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 496 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
431	G -> A	Yes
516	G -> C	Yes
644	G -> A	Yes
646	C -> A	Yes
754	C ->	No
754	C -> A	No
836	C -> T	Yes
891	C -> T	Yes
906	T ->	No
972	G ->	No
972	G -> C	No
1001	A ->	No
1031	C ->	No
1154	C ->	No
1315	C -> T	Yes
1375	G ->	No
1406	C -> T	Yes
1427	C ->	No
1497	T -> A	No
1508	C -> T	Yes
1543	G -> C	Yes
1600	T -> A	Yes
1728	C -> A	Yes
1747	C -> T	No

1827	C -> T	Yes
1843	A -> G	No
1843	A -> T	No
1908	A -> G	Yes
1932	T -> A	No
2009	G -> T	Yes
2142	G -> A	Yes
2171	G -> T	Yes
2351	G -> A	Yes
2547	C -> T	Yes
2603	G -> T	Yes
2743	C -> T	Yes
2844	C -> T	Yes
3159	A -> G	No

Variant protein HUMPHOSLIP_PEA_2_P35 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMPHOSLIP_PEA_2_T18. An alignment is given to the known protein (Phospholipid transfer protein precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMPHOSLIP_PEA_2_P35 and PLTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMPHOSLIP_PEA_2_P35, comprising a first amino acid sequence being at least 90 % homologous to MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGH FYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLRFRRQLLYWF corresponding to amino acids 1 - 109 of PLTP_HUMAN, which also corresponds to amino acids 1 - 109 of HUMPHOSLIP_PEA_2_P35, a second amino acid sequence bridging amino acid sequence comprising of L, a third amino acid sequence being at least 90 % homologous to

KVYDFLSTFITSGMRFLLNQQ corresponding to amino acids 163 - 183 of PLTP_HUMAN, which also corresponds to amino acids 111 - 131 of HUMPHOSLIP_PEA_2_P35, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VWAATGRRVARVGMLSL corresponding to amino acids 132 - 148 of HUMPHOSLIP_PEA_2_P35, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for an edge portion of HUMPHOSLIP_PEA_2_P35, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise FLK having a structure as follows (numbering according to HUMPHOSLIP_PEA_2_P35): a sequence starting from any of amino acid numbers 109-x to 109; and ending at any of amino acid numbers 111 + ((n-2) - x), in which x varies from 0 to n-2.

3.An isolated polypeptide encoding for a tail of HUMPHOSLIP_PEA_2_P35, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VWAATGRRVARVGMLSL in HUMPHOSLIP_PEA_2_P35.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMPHOSLIP_PEA_2_P35 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 497, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein

HUMPHOSLIP_PEA_2_P35 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 497 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	H -> R	Yes
18	E -> V	Yes
81	D -> H	Yes

The glycosylation sites of variant protein HUMPHOSLIP_PEA_2_P35, as compared to the known protein Phospholipid transfer protein precursor, are described in Table 498 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 498 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
94	yes	94
143	no	
64	yes	64
245	no	
398	no	
117	no	

Variant protein HUMPHOSLIP_PEA_2_P35 is encoded by the following transcript(s): HUMPHOSLIP_PEA_2_T18, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMPHOSLIP_PEA_2_T18 is shown in bold; this coding portion starts at position 276 and ends at position 719. The transcript also has the following SNPs as listed in Table 499 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is

known or not; the presence of known SNPs in variant protein HUMPHOSLIP_PEA_2_P35 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 499 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
174	G -> T	No
175	A -> T	No
322	A -> G	Yes
328	A -> T	Yes
431	G -> A	Yes
516	G -> C	Yes
765	C -> T	Yes
827	C -> T	Yes
842	T ->	No
908	G ->	No
908	G -> C	No
937	A ->	No
967	C ->	No
1090	C ->	No
1251	C -> T	Yes
1311	G ->	No
1342	C -> T	Yes
1363	C ->	No
1433	T -> A	No
1444	C -> T	Yes
1479	G -> C	Yes
1536	T -> A	Yes
1664	C -> A	Yes

1683	C -> T	No
1763	C -> T	Yes
1779	A -> G	No
1779	A -> T	No
1844	A -> G	Yes
1868	T -> A	No
1945	G -> T	Yes
2078	G -> A	Yes
2107	G -> T	Yes
2287	G -> A	Yes
2483	C -> T	Yes
2539	G -> T	Yes
2679	C -> T	Yes
2780	C -> T	Yes
3095	A -> G	No

As noted above, cluster HUMPHOSLIP features 53 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMPHOSLIP_PEA_2_node_0 according to the present invention is supported by 150 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 500 below describes the starting and ending position of this segment on each transcript.

Table 500 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMPHOSLIP_PEA_2_T6	1	264
HUMPHOSLIP_PEA_2_T7	1	264
HUMPHOSLIP_PEA_2_T14	1	264
HUMPHOSLIP_PEA_2_T16	1	264
HUMPHOSLIP_PEA_2_T17	1	264
HUMPHOSLIP_PEA_2_T18	1	264
HUMPHOSLIP_PEA_2_T19	1	264

Segment cluster HUMPHOSLIP_PEA_2_node_19 according to the present invention is supported by 186 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16 and HUMPHOSLIP_PEA_2_T19. Table 501 below describes the starting and ending position of this segment on each transcript.

Table 501 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	559	714
HUMPHOSLIP_PEA_2_T7	697	852
HUMPHOSLIP_PEA_2_T14	605	760
HUMPHOSLIP_PEA_2_T16	605	760
HUMPHOSLIP_PEA_2_T19	605	760

Segment cluster HUMPHOSLIP_PEA_2_node_34 according to the present invention is supported by 191 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16,

HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 502 below describes the starting and ending position of this segment on each transcript.

Table 502 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	971	1111
HUMPHOSLIP_PEA_2_T7	1109	1249
HUMPHOSLIP_PEA_2_T14	1102	1242
HUMPHOSLIP_PEA_2_T16	1010	1150
HUMPHOSLIP_PEA_2_T17	732	872
HUMPHOSLIP_PEA_2_T18	946	1086
HUMPHOSLIP_PEA_2_T19	1017	1157

Segment cluster HUMPHOSLIP_PEA_2_node_68 according to the present invention is supported by 131 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 503 below describes the starting and ending position of this segment on each transcript.

Table 503 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1867	2285
HUMPHOSLIP_PEA_2_T7	2005	2423
HUMPHOSLIP_PEA_2_T14	1998	2416
HUMPHOSLIP_PEA_2_T16	1906	2324
HUMPHOSLIP_PEA_2_T17	1628	2046
HUMPHOSLIP_PEA_2_T18	1842	2260

HUMPHOSLIP_PEA_2_T19	1996	2414

Segment cluster HUMPHOSLIP_PEA_2_node_70 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 504 below describes the starting and ending position of this segment on each transcript.

Table 504 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2298	2529
HUMPHOSLIP_PEA_2_T7	2436	2667
HUMPHOSLIP_PEA_2_T14	2429	2660
HUMPHOSLIP_PEA_2_T16	2337	2568
HUMPHOSLIP_PEA_2_T17	2059	2290
HUMPHOSLIP_PEA_2_T18	2273	2504
HUMPHOSLIP_PEA_2_T19	2427	2658

Segment cluster HUMPHOSLIP_PEA_2_node_75 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 505 below describes the starting and ending position of this segment on each transcript.

Table 505 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2846	3125
HUMPHOSLIP_PEA_2_T7	2984	3263
HUMPHOSLIP_PEA_2_T14	2977	3256
HUMPHOSLIP_PEA_2_T16	2885	3164
HUMPHOSLIP_PEA_2_T17	2607	2886
HUMPHOSLIP_PEA_2_T18	2821	3100
HUMPHOSLIP_PEA_2_T19	2975	3254

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMPHOSLIP_PEA_2_node_2 according to the present invention is supported by 159 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 506 below describes the starting and ending position of this segment on each transcript.

Table 506 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	265	337
HUMPHOSLIP_PEA_2_T7	265	337
HUMPHOSLIP_PEA_2_T14	265	337
HUMPHOSLIP_PEA_2_T16	265	337

HUMPHOSLIP_PEA_2_T17	265	337
HUMPHOSLIP_PEA_2_T18	265	337
HUMPHOSLIP_PEA_2_T19	265	337

Segment cluster HUMPHOSLIP_PEA_2_node_3 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 507 below describes the starting and ending position of this segment on each transcript.

Table 507 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T7	338	355
HUMPHOSLIP_PEA_2_T14	338	355
HUMPHOSLIP_PEA_2_T16	338	355
HUMPHOSLIP_PEA_2_T17	338	355
HUMPHOSLIP_PEA_2_T18	338	355
HUMPHOSLIP_PEA_2_T19	338	355

Segment cluster HUMPHOSLIP_PEA_2_node_4 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 508 below describes the starting and ending position of this segment on each transcript.

Table 508 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMPHOSLIP_PEA_2_T7	356	375
HUMPHOSLIP_PEA_2_T14	356	375
HUMPHOSLIP_PEA_2_T16	356	375
HUMPHOSLIP_PEA_2_T17	356	375
HUMPHOSLIP_PEA_2_T18	356	375
HUMPHOSLIP_PEA_2_T19	356	375

Segment cluster HUMPHOSLIP_PEA_2_node_6 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 509 below describes the starting and ending position of this segment on each transcript.

Table 509 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T7	376	383
HUMPHOSLIP_PEA_2_T14	376	383
HUMPHOSLIP_PEA_2_T16	376	383
HUMPHOSLIP_PEA_2_T17	376	383
HUMPHOSLIP_PEA_2_T18	376	383
HUMPHOSLIP_PEA_2_T19	376	383

Segment cluster HUMPHOSLIP_PEA_2_node_7 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 510 below describes the starting and ending position of this segment on each transcript.

Table 510 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	338	343
HUMPHOSLIP_PEA_2_T7	384	389
HUMPHOSLIP_PEA_2_T14	384	389
HUMPHOSLIP_PEA_2_T16	384	389
HUMPHOSLIP_PEA_2_T17	384	389
HUMPHOSLIP_PEA_2_T18	384	389
HUMPHOSLIP_PEA_2_T19	384	389

Segment cluster HUMPHOSLIP_PEA_2_node_8 according to the present invention is supported by 171 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 511 below describes the starting and ending position of this segment on each transcript.

Table 511 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	344	378
HUMPHOSLIP_PEA_2_T7	390	424
HUMPHOSLIP_PEA_2_T14	390	424
HUMPHOSLIP_PEA_2_T16	390	424
HUMPHOSLIP_PEA_2_T17	390	424
HUMPHOSLIP_PEA_2_T18	390	424
HUMPHOSLIP_PEA_2_T19	390	424

Segment cluster HUMPHOSLIP_PEA_2_node_9 according to the present invention is supported by 168 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 512 below describes the starting and ending position of this segment on each transcript.

Table 512 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	379	429
HUMPHOSLIP_PEA_2_T7	425	475
HUMPHOSLIP_PEA_2_T14	425	475
HUMPHOSLIP_PEA_2_T16	425	475
HUMPHOSLIP_PEA_2_T17	425	475
HUMPHOSLIP_PEA_2_T18	425	475
HUMPHOSLIP_PEA_2_T19	425	475

Segment cluster HUMPHOSLIP_PEA_2_node_14 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T7. Table 513 below describes the starting and ending position of this segment on each transcript.

Table 513 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T7	476	567

Segment cluster HUMPHOSLIP_PEA_2_node_15 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6,

HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 514 below describes the starting and ending position of this segment on each transcript.

Table 514 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	430	445
HUMPHOSLIP_PEA_2_T7	568	583
HUMPHOSLIP_PEA_2_T14	476	491
HUMPHOSLIP_PEA_2_T16	476	491
HUMPHOSLIP_PEA_2_T18	476	491
HUMPHOSLIP_PEA_2_T19	476	491

Segment cluster HUMPHOSLIP_PEA_2_node_16 according to the present invention is supported by 179 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 515 below describes the starting and ending position of this segment on each transcript.

Table 515 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	446	534
HUMPHOSLIP_PEA_2_T7	584	672
HUMPHOSLIP_PEA_2_T14	492	580
HUMPHOSLIP_PEA_2_T16	492	580
HUMPHOSLIP_PEA_2_T18	492	580
HUMPHOSLIP_PEA_2_T19	492	580

Segment cluster HUMPHOSLIP_PEA_2_node_17 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 516 below describes the starting and ending position of this segment on each transcript.

Table 516 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	535	558
HUMPHOSLIP_PEA_2_T7	673	696
HUMPHOSLIP_PEA_2_T14	581	604
HUMPHOSLIP_PEA_2_T16	581	604
HUMPHOSLIP_PEA_2_T18	581	604
HUMPHOSLIP_PEA_2_T19	581	604

Segment cluster HUMPHOSLIP_PEA_2_node_23 according to the present invention is supported by 168 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 517 below describes the starting and ending position of this segment on each transcript.

Table 517 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	715	766
HUMPHOSLIP_PEA_2_T7	853	904

HUMPHOSLIP_PEA_2_T14	761	812
HUMPHOSLIP_PEA_2_T16	761	812
HUMPHOSLIP_PEA_2_T17	476	527
HUMPHOSLIP_PEA_2_T18	605	656
HUMPHOSLIP_PEA_2_T19	761	812

Segment cluster HUMPHOSLIP_PEA_2_node_24 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 518 below describes the starting and ending position of this segment on each transcript.

Table 518 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	767	778
HUMPHOSLIP_PEA_2_T7	905	916
HUMPHOSLIP_PEA_2_T14	813	824
HUMPHOSLIP_PEA_2_T16	813	824
HUMPHOSLIP_PEA_2_T17	528	539
HUMPHOSLIP_PEA_2_T18	657	668
HUMPHOSLIP_PEA_2_T19	813	824

Segment cluster HUMPHOSLIP_PEA_2_node_25 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T14 and HUMPHOSLIP_PEA_2_T18. Table 519 below describes the starting and ending position of this segment on each transcript.

Table 519- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T14	825	909
HUMPHOSLIP_PEA_2_T18	669	753

Segment cluster HUMPHOSLIP_PEA_2_node_26 according to the present invention is supported by 163 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 520 below describes the starting and ending position of this segment on each transcript.

Table 520 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	779	842
HUMPHOSLIP_PEA_2_T7	917	980
HUMPHOSLIP_PEA_2_T14	910	973
HUMPHOSLIP_PEA_2_T16	825	888
HUMPHOSLIP_PEA_2_T17	540	603
HUMPHOSLIP_PEA_2_T18	754	817
HUMPHOSLIP_PEA_2_T19	825	888

Segment cluster HUMPHOSLIP_PEA_2_node_29 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 521 below describes the starting and ending position of this segment on each transcript.

Table 521 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	843	849
HUMPHOSLIP_PEA_2_T7	981	987
HUMPHOSLIP_PEA_2_T14	974	980
HUMPHOSLIP_PEA_2_T17	604	610
HUMPHOSLIP_PEA_2_T18	818	824
HUMPHOSLIP_PEA_2_T19	889	895

Segment cluster HUMPHOSLIP_PEA_2_node_30 according to the present invention is supported by 181 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 522 below describes the starting and ending position of this segment on each transcript.

Table 522- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	850	934
HUMPHOSLIP_PEA_2_T7	988	1072
HUMPHOSLIP_PEA_2_T14	981	1065
HUMPHOSLIP_PEA_2_T16	889	973
HUMPHOSLIP_PEA_2_T17	611	695
HUMPHOSLIP_PEA_2_T18	825	909
HUMPHOSLIP_PEA_2_T19	896	980

Segment cluster HUMPHOSLIP_PEA_2_node_33 according to the present invention is supported by 173 libraries. The number of libraries was determined as previously described.

This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 523 below describes the starting and ending position of this segment on each transcript.

Table 523 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	935	970
HUMPHOSLIP_PEA_2_T7	1073	1108
HUMPHOSLIP_PEA_2_T14	1066	1101
HUMPHOSLIP_PEA_2_T16	974	1009
HUMPHOSLIP_PEA_2_T17	696	731
HUMPHOSLIP_PEA_2_T18	910	945
HUMPHOSLIP_PEA_2_T19	981	1016

Segment cluster HUMPHOSLIP_PEA_2_node_36 according to the present invention is supported by 163 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 524 below describes the starting and ending position of this segment on each transcript.

Table 524- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1112	1156
HUMPHOSLIP_PEA_2_T7	1250	1294
HUMPHOSLIP_PEA_2_T14	1243	1287
HUMPHOSLIP_PEA_2_T16	1151	1195

HUMPHOSLIP_PEA_2_T17	873	917
HUMPHOSLIP_PEA_2_T18	1087	1131
HUMPHOSLIP_PEA_2_T19	1158	1202

Segment cluster HUMPHOSLIP_PEA_2_node_37 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 525 below describes the starting and ending position of this segment on each transcript.

Table 525 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1157	1171
HUMPHOSLIP_PEA_2_T7	1295	1309
HUMPHOSLIP_PEA_2_T14	1288	1302
HUMPHOSLIP_PEA_2_T16	1196	1210
HUMPHOSLIP_PEA_2_T17	918	932
HUMPHOSLIP_PEA_2_T18	1132	1146
HUMPHOSLIP_PEA_2_T19	1203	1217

Segment cluster HUMPHOSLIP_PEA_2_node_39 according to the present invention is supported by 166 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 525 below describes the starting and ending position of this segment on each transcript.

Table 525 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1172	1201
HUMPHOSLIP_PEA_2_T7	1310	1339
HUMPHOSLIP_PEA_2_T14	1303	1332
HUMPHOSLIP_PEA_2_T16	1211	1240
HUMPHOSLIP_PEA_2_T17	933	962
HUMPHOSLIP_PEA_2_T18	1147	1176
HUMPHOSLIP_PEA_2_T19	1218	1247

Segment cluster HUMPHOSLIP_PEA_2_node_40 according to the present invention is supported by 199 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 526 below describes the starting and ending position of this segment on each transcript.

Table 526 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1202	1288
HUMPHOSLIP_PEA_2_T7	1340	1426
HUMPHOSLIP_PEA_2_T14	1333	1419
HUMPHOSLIP_PEA_2_T16	1241	1327
HUMPHOSLIP_PEA_2_T17	963	1049
HUMPHOSLIP_PEA_2_T18	1177	1263
HUMPHOSLIP_PEA_2_T19	1248	1334

Segment cluster HUMPHOSLIP_PEA_2_node_41 according to the present invention is supported by 186 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 527 below describes the starting and ending position of this segment on each transcript.

Table 527 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1289	1318
HUMPHOSLIP_PEA_2_T7	1427	1456
HUMPHOSLIP_PEA_2_T14	1420	1449
HUMPHOSLIP_PEA_2_T16	1328	1357
HUMPHOSLIP_PEA_2_T17	1050	1079
HUMPHOSLIP_PEA_2_T18	1264	1293
HUMPHOSLIP_PEA_2_T19	1335	1364

Segment cluster HUMPHOSLIP_PEA_2_node_42 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 528 below describes the starting and ending position of this segment on each transcript.

Table 528 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1319	1336
HUMPHOSLIP_PEA_2_T7	1457	1474
HUMPHOSLIP_PEA_2_T14	1450	1467

HUMPHOSLIP_PEA_2_T16	1358	1375
HUMPHOSLIP_PEA_2_T17	1080	1097
HUMPHOSLIP_PEA_2_T18	1294	1311
HUMPHOSLIP_PEA_2_T19	1365	1382

Segment cluster HUMPHOSLIP_PEA_2_node_44 according to the present invention is supported by 185 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 529 below describes the starting and ending position of this segment on each transcript.

Table 529 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1337	1363
HUMPHOSLIP_PEA_2_T7	1475	1501
HUMPHOSLIP_PEA_2_T14	1468	1494
HUMPHOSLIP_PEA_2_T16	1376	1402
HUMPHOSLIP_PEA_2_T17	1098	1124
HUMPHOSLIP_PEA_2_T18	1312	1338
HUMPHOSLIP_PEA_2_T19	1383	1409

Segment cluster HUMPHOSLIP_PEA_2_node_45 according to the present invention is supported by 197 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 530 below describes the starting and ending position of this segment on each transcript.

Table 530 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1364	1404
HUMPHOSLIP_PEA_2_T7	1502	1542
HUMPHOSLIP_PEA_2_T14	1495	1535
HUMPHOSLIP_PEA_2_T16	1403	1443
HUMPHOSLIP_PEA_2_T17	1125	1165
HUMPHOSLIP_PEA_2_T18	1339	1379
HUMPHOSLIP_PEA_2_T19	1410	1450

Segment cluster HUMPHOSLIP_PEA_2_node_47 according to the present invention is supported by 223 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 531 below describes the starting and ending position of this segment on each transcript.

Table 531 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1405	1447
HUMPHOSLIP_PEA_2_T7	1543	1585
HUMPHOSLIP_PEA_2_T14	1536	1578
HUMPHOSLIP_PEA_2_T16	1444	1486
HUMPHOSLIP_PEA_2_T17	1166	1208
HUMPHOSLIP_PEA_2_T18	1380	1422
HUMPHOSLIP_PEA_2_T19	1451	1493

Segment cluster HUMPHOSLIP_PEA_2_node_51 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 532 below describes the starting and ending position of this segment on each transcript.

Table 532 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1448	1462
HUMPHOSLIP_PEA_2_T7	1586	1600
HUMPHOSLIP_PEA_2_T14	1579	1593
HUMPHOSLIP_PEA_2_T16	1487	1501
HUMPHOSLIP_PEA_2_T17	1209	1223
HUMPHOSLIP_PEA_2_T18	1423	1437
HUMPHOSLIP_PEA_2_T19	1494	1508

Segment cluster HUMPHOSLIP_PEA_2_node_52 according to the present invention is supported by 235 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 533 below describes the starting and ending position of this segment on each transcript.

Table 533 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1463	1511
HUMPHOSLIP_PEA_2_T7	1601	1649
HUMPHOSLIP_PEA_2_T14	1594	1642

HUMPHOSLIP_PEA_2_T16	1502	1550
HUMPHOSLIP_PEA_2_T17	1224	1272
HUMPHOSLIP_PEA_2_T18	1438	1486
HUMPHOSLIP_PEA_2_T19	1509	1557

Segment cluster HUMPHOSLIP_PEA_2_node_53 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T19. Table 534 below describes the starting and ending position of this segment on each transcript.

Table 534 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T19	1558	1640

Segment cluster HUMPHOSLIP_PEA_2_node_54 according to the present invention is supported by 236 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 535 below describes the starting and ending position of this segment on each transcript.

Table 535 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1512	1552
HUMPHOSLIP_PEA_2_T7	1650	1690
HUMPHOSLIP_PEA_2_T14	1643	1683
HUMPHOSLIP_PEA_2_T16	1551	1591

HUMPHOSLIP_PEA_2_T17	1273	1313
HUMPHOSLIP_PEA_2_T18	1487	1527
HUMPHOSLIP_PEA_2_T19	1641	1681

Segment cluster HUMPHOSLIP_PEA_2_node_55 according to the present invention is supported by 232 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 536 below describes the starting and ending position of this segment on each transcript.

Table 536 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1553	1588
HUMPHOSLIP_PEA_2_T7	1691	1726
HUMPHOSLIP_PEA_2_T14	1684	1719
HUMPHOSLIP_PEA_2_T16	1592	1627
HUMPHOSLIP_PEA_2_T17	1314	1349
HUMPHOSLIP_PEA_2_T18	1528	1563
HUMPHOSLIP_PEA_2_T19	1682	1717

Segment cluster HUMPHOSLIP_PEA_2_node_58 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 537 below describes the starting and ending position of this segment on each transcript.

Table 537 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1589	1612
HUMPHOSLIP_PEA_2_T7	1727	1750
HUMPHOSLIP_PEA_2_T14	1720	1743
HUMPHOSLIP_PEA_2_T16	1628	1651
HUMPHOSLIP_PEA_2_T17	1350	1373
HUMPHOSLIP_PEA_2_T18	1564	1587
HUMPHOSLIP_PEA_2_T19	1718	1741

Segment cluster HUMPHOSLIP_PEA_2_node_59 according to the present invention is supported by 230 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 538 below describes the starting and ending position of this segment on each transcript.

Table 538 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1613	1648
HUMPHOSLIP_PEA_2_T7	1751	1786
HUMPHOSLIP_PEA_2_T14	1744	1779
HUMPHOSLIP_PEA_2_T16	1652	1687
HUMPHOSLIP_PEA_2_T17	1374	1409
HUMPHOSLIP_PEA_2_T18	1588	1623
HUMPHOSLIP_PEA_2_T19	1742	1777

Segment cluster HUMPHOSLIP_PEA_2_node_60 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 539 below describes the starting and ending position of this segment on each transcript.

Table 539 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1649	1671
HUMPHOSLIP_PEA_2_T7	1787	1809
HUMPHOSLIP_PEA_2_T14	1780	1802
HUMPHOSLIP_PEA_2_T16	1688	1710
HUMPHOSLIP_PEA_2_T17	1410	1432
HUMPHOSLIP_PEA_2_T18	1624	1646
HUMPHOSLIP_PEA_2_T19	1778	1800

Segment cluster HUMPHOSLIP_PEA_2_node_61 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 540 below describes the starting and ending position of this segment on each transcript.

Table 540 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1672	1680
HUMPHOSLIP_PEA_2_T7	1810	1818
HUMPHOSLIP_PEA_2_T14	1803	1811
HUMPHOSLIP_PEA_2_T16	1711	1719

HUMPHOSLIP_PEA_2_T17	1433	1441
HUMPHOSLIP_PEA_2_T18	1647	1655
HUMPHOSLIP_PEA_2_T19	1801	1809

Segment cluster HUMPHOSLIP_PEA_2_node_62 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 541 below describes the starting and ending position of this segment on each transcript.

Table 541 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1681	1703
HUMPHOSLIP_PEA_2_T7	1819	1841
HUMPHOSLIP_PEA_2_T14	1812	1834
HUMPHOSLIP_PEA_2_T16	1720	1742
HUMPHOSLIP_PEA_2_T17	1442	1464
HUMPHOSLIP_PEA_2_T18	1656	1678
HUMPHOSLIP_PEA_2_T19	1810	1832

Segment cluster HUMPHOSLIP_PEA_2_node_63 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 542 below describes the starting and ending position of this segment on each transcript.

Table 542 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1704	1727
HUMPHOSLIP_PEA_2_T7	1842	1865
HUMPHOSLIP_PEA_2_T14	1835	1858
HUMPHOSLIP_PEA_2_T16	1743	1766
HUMPHOSLIP_PEA_2_T17	1465	1488
HUMPHOSLIP_PEA_2_T18	1679	1702
HUMPHOSLIP_PEA_2_T19	1833	1856

Segment cluster HUMPHOSLIP_PEA_2_node_64 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 543 below describes the starting and ending position of this segment on each transcript.

Table 543 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1728	1734
HUMPHOSLIP_PEA_2_T7	1866	1872
HUMPHOSLIP_PEA_2_T14	1859	1865
HUMPHOSLIP_PEA_2_T16	1767	1773
HUMPHOSLIP_PEA_2_T17	1489	1495
HUMPHOSLIP_PEA_2_T18	1703	1709
HUMPHOSLIP_PEA_2_T19	1857	1863

Segment cluster HUMPHOSLIP_PEA_2_node_65 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6,

HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 544 below describes the starting and ending position of this segment on each transcript.

Table 544 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1735	1754
HUMPHOSLIP_PEA_2_T7	1873	1892
HUMPHOSLIP_PEA_2_T14	1866	1885
HUMPHOSLIP_PEA_2_T16	1774	1793
HUMPHOSLIP_PEA_2_T17	1496	1515
HUMPHOSLIP_PEA_2_T18	1710	1729
HUMPHOSLIP_PEA_2_T19	1864	1883

Segment cluster HUMPHOSLIP_PEA_2_node_66 according to the present invention is supported by 180 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 545 below describes the starting and ending position of this segment on each transcript.

Table 545 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1755	1844
HUMPHOSLIP_PEA_2_T7	1893	1982
HUMPHOSLIP_PEA_2_T14	1886	1975
HUMPHOSLIP_PEA_2_T16	1794	1883
HUMPHOSLIP_PEA_2_T17	1516	1605

HUMPHOSLIP_PEA_2_T18	1730	1819
HUMPHOSLIP_PEA_2_T19	1884	1973

Segment cluster HUMPHOSLIP_PEA_2_node_67 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 546 below describes the starting and ending position of this segment on each transcript.

Table 546 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	1845	1866
HUMPHOSLIP_PEA_2_T7	1983	2004
HUMPHOSLIP_PEA_2_T14	1976	1997
HUMPHOSLIP_PEA_2_T16	1884	1905
HUMPHOSLIP_PEA_2_T17	1606	1627
HUMPHOSLIP_PEA_2_T18	1820	1841
HUMPHOSLIP_PEA_2_T19	1974	1995

Segment cluster HUMPHOSLIP_PEA_2_node_69 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 547 below describes the starting and ending position of this segment on each transcript.

Table 547 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMPHOSLIP_PEA_2_T6	2286	2297
HUMPHOSLIP_PEA_2_T7	2424	2435
HUMPHOSLIP_PEA_2_T14	2417	2428
HUMPHOSLIP_PEA_2_T16	2325	2336
HUMPHOSLIP_PEA_2_T17	2047	2058
HUMPHOSLIP_PEA_2_T18	2261	2272
HUMPHOSLIP_PEA_2_T19	2415	2426

Segment cluster HUMPHOSLIP_PEA_2_node_71 according to the present invention can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 548 below describes the starting and ending position of this segment on each transcript.

Table 548 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2530	2542
HUMPHOSLIP_PEA_2_T7	2668	2680
HUMPHOSLIP_PEA_2_T14	2661	2673
HUMPHOSLIP_PEA_2_T16	2569	2581
HUMPHOSLIP_PEA_2_T17	2291	2303
HUMPHOSLIP_PEA_2_T18	2505	2517
HUMPHOSLIP_PEA_2_T19	2659	2671

Segment cluster HUMPHOSLIP_PEA_2_node_72 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16,

HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 549 below describes the starting and ending position of this segment on each transcript.

Table 549 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2543	2647
HUMPHOSLIP_PEA_2_T7	2681	2785
HUMPHOSLIP_PEA_2_T14	2674	2778
HUMPHOSLIP_PEA_2_T16	2582	2686
HUMPHOSLIP_PEA_2_T17	2304	2408
HUMPHOSLIP_PEA_2_T18	2518	2622
HUMPHOSLIP_PEA_2_T19	2672	2776

Segment cluster HUMPHOSLIP_PEA_2_node_73 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 550 below describes the starting and ending position of this segment on each transcript.

Table 550 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2648	2755
HUMPHOSLIP_PEA_2_T7	2786	2893
HUMPHOSLIP_PEA_2_T14	2779	2886
HUMPHOSLIP_PEA_2_T16	2687	2794
HUMPHOSLIP_PEA_2_T17	2409	2516
HUMPHOSLIP_PEA_2_T18	2623	2730

HUMPHOSLIP_PEA_2_T19	2777	2884

Segment cluster HUMPHOSLIP_PEA_2_node_74 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMPHOSLIP_PEA_2_T6, HUMPHOSLIP_PEA_2_T7, HUMPHOSLIP_PEA_2_T14, HUMPHOSLIP_PEA_2_T16, HUMPHOSLIP_PEA_2_T17, HUMPHOSLIP_PEA_2_T18 and HUMPHOSLIP_PEA_2_T19. Table 551below describes the starting and ending position of this segment on each transcript.

Table 551 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMPHOSLIP_PEA_2_T6	2756	2845
HUMPHOSLIP_PEA_2_T7	2894	2983
HUMPHOSLIP_PEA_2_T14	2887	2976
HUMPHOSLIP_PEA_2_T16	2795	2884
HUMPHOSLIP_PEA_2_T17	2517	2606
HUMPHOSLIP_PEA_2_T18	2731	2820
HUMPHOSLIP_PEA_2_T19	2885	2974

Variant protein alignment to the previously known protein:

Sequence name: PLTP_HUMAN

Sequence documentation:

```
Alignment of: HUMPHOSLIP_PEA_2_P10 x PLTP_HUMAN    ..

Alignment segment 1/1:

                    Quality: 3716.00
Escore:       0
              Matching length:     398                Total
length:     493
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:   80.73      Total Percent
Identity:   80.73
                         Gaps:       1

Alignment:
                 .         .         .         .         .
         1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
           ||||||||||||||||||||||||||||||||||||||||||||||||||
         1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
                 .         .         .         .         .
        51 IPDLRGKEGHFYYNISE..................................  67
           |||||||||||||||||
        51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
                 .         .         .         .         .
        67 ..................................................  67

       101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
                 .         .         .         .         .
        68 ............KVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 105
                      ||||||||||||||||||||||||||||||||||||
```

```
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 200
                .         .         .         .         .
106 DTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRGAFFPLTERNWSLPN 155
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 DTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRGAFFPLTERNWSLPN 250
                .         .         .         .         .
156 RAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLD 205
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 RAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLD 300
                .         .         .         .         .
206 MLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASV 255
    |||||||||||||||||||||||||||||||||||||||||||||||||||
301 MLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASV 350
                .         .         .         .         .
256 TIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHS 305
    |||||||||||||||||||||||||||||||||||||||||||||||||||
351 TIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHS 400
                .         .         .         .         .
306 ALESLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVT 355
    |||||||||||||||||||||||||||||||||||||||||||||||||||
401 ALESLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVT 450
                .         .         .         .
356 NHAGFLTIGADLHFAKGLREVIEKNRPADVRASTAPTPSTAAV         398
    |||||||||||||||||||||||||||||||||||||||||
451 NHAGFLTIGADLHFAKGLREVIEKNRPADVRASTAPTPSTAAV         493
```

Sequence name: PLTP_HUMAN

Sequence documentation:

Alignment of: HUMPHOSLIP_PEA_2_P12 x PLTP_HUMAN ..

Alignment segment 1/1:

Quality: 4101.00

Escore: 0

Matching length: 427 Total length: 427

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
                .         .         .         .         .
     51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
                .         .         .         .         .
    101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
```

```
                .         .         .         .         .
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 200
                .         .         .         .         .
201 DTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRGAFFPLTERNWSLPN 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 DTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRGAFFPLTERNWSLPN 250
                .         .         .         .         .
251 RAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLD 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 RAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLD 300
                .         .         .         .         .
301 MLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASV 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 MLLRATYFGSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASV 350
                .         .         .         .         .
351 TIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHS 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 TIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHS 400
                .         .
401 ALESLALIPLQAPLKTMLQIGVMPMLN                         427
    |||||||||||||||||||||||||||
401 ALESLALIPLQAPLKTMLQIGVMPMLN                         427
```

Sequence name: PLTP_HUMAN

Sequence documentation:

Alignment of: HUMPHOSLIP_PEA_2_P31 x PLTP_HUMAN ..

Alignment segment 1/1:

Quality: 639.00
Escore: 0
Matching length: 67 Total length: 67
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
                  .
       51 IPDLRGKEGHFYYNISE                                  67
          |||||||||||||||||
       51 IPDLRGKEGHFYYNISE                                  67
```

Sequence name: PLTP_HUMAN

Sequence documentation:

Alignment of: HUMPHOSLIP_PEA_2_P33 x PLTP_HUMAN ..

Alignment segment 1/1:

Quality: 1767.00

Escore: 0

Matching length: 184 Total length: 184

Matching Percent Similarity: 100.00 Matching Percent Identity: 99.46

Total Percent Similarity: 100.00 Total Percent Identity: 99.46

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
                .         .         .         .         .
     51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
                .         .         .         .         .
    101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
                .         .         .
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQV                 184
    |||||||||||||||||||||||||||||||:
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQI                 184
```

Sequence name: PLTP_HUMAN

Sequence documentation:

Alignment of: HUMPHOSLIP_PEA_2_P34 x PLTP_HUMAN ..

Alignment segment 1/1:

Quality: 1971.00

Escore: 0

Matching length: 205 Total length: 205

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

640

```
  1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
            .         .         .         .         .
 51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
            .         .         .         .         .
101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
            .         .         .         .         .
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLL 200

201 DTVPV                                              205
    |||||
201 DTVPV                                              205
```

Sequence name: PLTP_HUMAN

Sequence documentation:

Alignment of: HUMPHOSLIP_PEA_2_P35 x PLTP_HUMAN ..

```
Alignment segment 1/1:

                    Quality: 1158.00
Escore:       0
              Matching length:     132                Total
length:     184
 Matching Percent Similarity:  100.00   Matching Percent
Identity:   98.48
    Total Percent Similarity:   71.74      Total Percent
Identity:   70.65
                         Gaps:       1

Alignment:
                .         .         .         .         .
        1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
          |||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETIT 50
                .         .         .         .         .
       51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||
       51 IPDLRGKEGHFYYNISEVKVTELQLTSSELDFQPQQELMLQITNASLGLR 100
                .         .         .         .         .
      101 FRRQLLYWFL..................................... 110
          |||||||||:
      101 FRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASV 150
                .         .         .
      111 ............KVYDFLSTFITSGMRFLLNQQV               132
                      |||||||||||||||||||:
      151 SRMHAAFGGTFKKVYDFLSTFITSGMRFLLNQQI               184
```

DESCRIPTION FOR CLUSTER AI076020

Cluster AI076020 features 1 transcript(s) and 8 segment(s) of interest, the names for which are given in Tables 552 and 553, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 554.

Table 552 - Transcripts of interest

Transcript Name	Sequence ID No.
AI076020_T0	58

Table 553 - Segments of interest

Segment Name	Sequence ID No.
AI076020_node_0	571
AI076020_node_3	572
AI076020_node_8	573
AI076020_node_1	574
AI076020_node_4	575
AI076020_node_5	576
AI076020_node_6	577
AI076020_node_7	578

Table 554 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
AI076020_P1	1334	AI076020_T0

These sequences are variants of the known protein C1q-related factor precursor (SwissProt accession identifier C1RF_HUMAN), SEQ ID NO: 1434, referred to herein as the previously known protein.

The sequence for protein C1q-related factor precursor is given at the end of the application, as "C1q-related factor precursor amino acid sequence".

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: locomotory behavior, which are annotation(s) related to Biological Process.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster AI076020 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 31 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 31 and Table 555. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: brain malignant tumors and a mixture of malignant tumors from different tissues.

Table 555 - Normal tissue distribution

Name of Tissue	Number
bone	0
brain	9
epithelial	0
general	4
kidney	2
lung	0
ovary	0
pancreas	30
uterus	0

Table 556 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bone	3.3e-01	5.9e-02	4.0e-01	2.5	2.4e-01	3.0
brain	8.8e-04	2.2e-03	5.5e-11	14.2	4.6e-08	8.7
epithelial	2.6e-01	8.6e-02	2.8e-01	2.4	1.8e-02	4.5
general	2.1e-03	3.0e-04	2.0e-06	4.3	8.4e-06	3.5
kidney	5.5e-01	3.3e-01	3.4e-01	2.3	8.2e-02	3.3
lung	1	6.3e-01	1	1.0	3.8e-01	2.2
ovary	4.2e-01	4.5e-01	0.0e+00	0.0	0.0e+00	0.0
pancreas	6.0e-01	7.1e-01	8.9e-01	0.6	9.5e-01	0.5
uterus	1	4.0e-01	1	1.0	6.4e-01	1.5

As noted above, cluster AI076020 features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein C1q-related factor precursor. A description of each variant protein according to the present invention is now provided.

Variant protein AI076020_P1 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) AI076020_T0. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein AI076020_P1 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 557, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AI076020_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 557 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
36	P -> R	Yes
66	Q -> R	Yes
165	K -> R	Yes

Variant protein AI076020_P1 is encoded by the following transcript(s): AI076020_T0, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript AI076020_T0 is shown in bold; this coding portion starts at position 261 and ends at position 1034. The transcript also has the following SNPs as listed in Table 558(given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein AI076020_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 558 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
367	C -> G	Yes
457	A -> G	Yes
464	C -> A	Yes
754	A -> G	Yes
1265	C -> T	Yes
1384	C -> T	Yes

1402	G -> C	Yes
1452	T -> C	Yes

As noted above, cluster AI076020 features 8 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster AI076020_node_0 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 559 below describes the starting and ending position of this segment on each transcript.

Table 559 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1	774

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 560.

Table 560 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
AI076020_0_3_0	lung malignant tumors	LUN

Segment cluster AI076020_node_3 according to the present invention is supported by 30 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 561 below describes the starting and ending position of this segment on each transcript.

Table 561 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	858	1027

Segment cluster AI076020_node_8 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 562 below describes the starting and ending position of this segment on each transcript.

Table 562 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1359	1533

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster AI076020_node_1 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 563 below describes the starting and ending position of this segment on each transcript.

Table 563 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	775	857

Segment cluster AI076020_node_4 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 564 below describes the starting and ending position of this segment on each transcript.

Table 564 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1028	1129

Segment cluster AI076020_node_5 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 565 below describes the starting and ending position of this segment on each transcript.

Table 565 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1130	1244

Segment cluster AI076020_node_6 according to the present invention is supported by 32 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 566 below describes the starting and ending position of this segment on each transcript.

Table 566 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1245	1320

Segment cluster AI076020_node_7 according to the present invention is supported by 33 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): AI076020_T0. Table 567below describes the starting and ending position of this segment on each transcript.

Table 567 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
AI076020_T0	1321	1358

DESCRIPTION FOR CLUSTER T23580

Cluster T23580 features 1 transcript(s) and 5 segment(s) of interest, the names for which are given in Tables 568 and 569, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 570.

Table 568 - Transcripts of interest

Transcript Name	Sequence ID No.
T23580_T10	1626

Table 569 - Segments of interest

Segment Name	Sequence ID No.
T23580_node_17	579
T23580_node_18	580
T23580_node_21	581
T23580_node_19	582
T23580_node_20	583

Table 570 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
T23580_P5	1335	T23580_T10

These sequences are variants of the known protein Neuronal protein NP25 (SwissProt accession identifier TAG3_HUMAN; known also according to the synonyms Neuronal protein 22; NP22; Transgelin-3), SEQ ID NO: 1435, referred to herein as the previously known protein and also as NP25_HUMAN, which is the former SwissProt accession identifier.

The sequence for protein Neuronal protein NP25 is given at the end of the application, as "Neuronal protein NP25 amino acid sequence".

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: central nervous system development, which are annotation(s) related to Biological Process.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

For this cluster, at least one oligonucleotide was found to demonstrate overexpression of the cluster, although not of at least one transcript/segment as listed below. Microarray (chip) data is also available for this cluster as follows. Various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer, as previously described. The following oligonucleotides were found to hit this cluster but not other segments/transcripts below, shown in Table 571, with regard to lung cancer.

Table 571 - Oligonucleotides related to this cluster

Oligonucleotide name	Overexpressed in cancers	Chip reference
T23580_0_0_902	lung malignant tumors	LUN

As noted above, cluster T23580 features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Neuronal protein NP25. A description of each variant protein according to the present invention is now provided.

Variant protein T23580_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T23580_T10. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Signal peptide,NN:NO) predicts that this protein has a signal peptide.

Variant protein T23580_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 572, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T23580_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 572 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
129	V -> I	Yes

Variant protein T23580_P5 is encoded by the following transcript(s): T23580_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T23580_T10 is shown in bold; this coding portion starts at position 1066 and ends at position 1485. The transcript also has the following SNPs as listed in Table 573 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T23580_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 573 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
37	A -> C	Yes

320	G -> A	Yes
371	G -> T	Yes
372	G -> A	Yes
441	A -> G	Yes
699	G -> C	Yes
744	C -> G	Yes
862	G -> T	Yes
1450	G -> A	Yes

As noted above, cluster T23580 features 5 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster T23580_node_17 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T23580_T10. Table 574 below describes the starting and ending position of this segment on each transcript.

Table 574 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T23580_T10	1	1098

Segment cluster T23580_node_18 according to the present invention is supported by 102 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T23580_T10. Table 575 below describes the starting and ending position of this segment on each transcript.

Table 575 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T23580_T10	1099	1357

Segment cluster T23580_node_21 according to the present invention is supported by 79 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T23580_T10. Table 576 below describes the starting and ending position of this segment on each transcript.

Table 576 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T23580_T10	1382	1582

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster T23580_node_19 according to the present invention can be found in the following transcript(s): T23580_T10. Table 577 below describes the starting and ending position of this segment on each transcript.

Table 577 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T23580_T10	1358	1370

Segment cluster T23580_node_20 according to the present invention can be found in the following transcript(s): T23580_T10. Table 578 below describes the starting and ending position of this segment on each transcript.

Table 578 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T23580_T10	1371	1381

DESCRIPTION FOR CLUSTER M79217

Cluster M79217 features 6 transcript(s) and 32 segment(s) of interest, the names for which are given in Tables 579 and 580, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 581.

Table 579 - Transcripts of interest

Transcript Name	Sequence ID No.
M79217_PEA_1_T1	59
M79217_PEA_1_T3	60
M79217_PEA_1_T8	61
M79217_PEA_1_T10	62
M79217_PEA_1_T15	63
M79217_PEA_1_T18	64

Table 580 - Segments of interest

Segment Name	Sequence ID No.
M79217_PEA_1_node_2	584
M79217_PEA_1_node_4	585
M79217_PEA_1_node_9	586
M79217_PEA_1_node_10	587
M79217_PEA_1_node_11	588
M79217_PEA_1_node_13	589
M79217_PEA_1_node_14	590

M79217_PEA_1_node_16	591
M79217_PEA_1_node_23	592
M79217_PEA_1_node_24	593
M79217_PEA_1_node_31	594
M79217_PEA_1_node_33	595
M79217_PEA_1_node_34	596
M79217_PEA_1_node_35	597
M79217_PEA_1_node_37	598
M79217_PEA_1_node_38	599
M79217_PEA_1_node_41	600
M79217_PEA_1_node_44	601
M79217_PEA_1_node_0	602
M79217_PEA_1_node_7	603
M79217_PEA_1_node_12	604
M79217_PEA_1_node_19	605
M79217_PEA_1_node_21	606
M79217_PEA_1_node_26	607
M79217_PEA_1_node_27	608
M79217_PEA_1_node_30	609
M79217_PEA_1_node_32	610
M79217_PEA_1_node_36	611
M79217_PEA_1_node_39	612
M79217_PEA_1_node_40	613
M79217_PEA_1_node_42	614
M79217_PEA_1_node_43	615

Table 581 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
M79217_PEA_1_P1	1336	M79217_PEA_1_T1;

		M79217_PEA_1_T3
M79217_PEA_1_P2	1337	M79217_PEA_1_T8
M79217_PEA_1_P4	1338	M79217_PEA_1_T10
M79217_PEA_1_P8	1339	M79217_PEA_1_T15
M79217_PEA_1_P11	1340	M79217_PEA_1_T18

These sequences are variants of the known protein Exostosin-like 3 (SwissProt accession identifier EXL3_HUMAN; known also according to the synonyms EC 2.4.1.223; Glucuronyl-galactosyl-proteoglycan 4- alpha-N-acetylglucosaminyltransferase; Putative tumor suppressor protein EXTL3; Multiple exostosis-like protein 3; Hereditary multiple exostoses gene isolog; EXT-related protein 1), SEQ ID NO: 1436, referred to herein as the previously known protein.

Protein Exostosin-like 3 is known or believed to have the following function(s): Probable glycosyltransferase (By similarity). The sequence for protein Exostosin-like 3 is given at the end of the application, as "Exostosin-like 3 amino acid sequence". Protein Exostosin-like 3 localization is believed to be Type II membrane protein. Endoplasmic reticulum.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: cell growth and/or maintenance, which are annotation(s) related to Biological Process; transferase, transferring glycosyl groups, which are annotation(s) related to Molecular Function; and endoplasmic reticulum; integral membrane protein, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster M79217 features 6 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Exostosin-like 3. A description of each variant protein according to the present invention is now provided.

Variant protein M79217_PEA_1_P1 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M79217_PEA_1_T1. An alignment is given to the known protein (Exostosin-like 3) at the end of the application. One or more alignments to one or more previously published protein

sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M79217_PEA_1_P1 and BAA25445 (SEQ ID NO: 1437):

1.An isolated chimeric polypeptide encoding for M79217_PEA_1_P1, comprising a first amino acid sequence being at least 90 % homologous to MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC RLHNCFDYSRCPLTSGFPVYYVYDSDQFVFGSYLDPLVKQAFQATARANVYYVTENADIA CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARD RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRD MVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFHERHKCINFF VKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 13 - 931 of BAA25445, which also corresponds to amino acids 1 - 919 of M79217_PEA_1_P1.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because the Signalp_hmm software predicts that this protein has a signal anchor region.

Variant protein M79217_PEA_1_P1 is encoded by the following transcript(s): M79217_PEA_1_T1, for which the sequence(s) is/are given at the end of the application. The

coding portion of transcript M79217_PEA_1_T1 is shown in bold; this coding portion starts at position 1074 and ends at position 3830. The transcript also has the following SNPs as listed in Table 582 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 582 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
1014	C -> T	No
1015	T ->	No
1072	T -> C	No
1232	T -> A	No
1383	A -> G	No
1440	A -> G	No
1544	C ->	No
1546	G -> A	No
1685	T -> G	No
2215	C ->	No
2300	A -> G	Yes
2483	T -> C	No
2518	C ->	No
2632	T -> G	No
3190	T -> C	Yes
3352	T -> C	No
3373	G -> T	No
3386	C ->	No
3449	C -> T	Yes
3618	A -> G	No
3733	A -> G	No

4021	C ->	No
4021	C -> T	No
4086	G -> A	No
4087	G -> A	No
4416	T -> A	No
4586	G -> A	Yes
4772	C -> T	No
5110	C -> T	Yes
5219	C -> T	Yes
5437	G -> A	No
5645	G -> A	No
5743	G -> A	Yes
5887	G -> T	Yes
6143	A -> C	No
6277	G ->	No
6277	G -> C	No
6295	C -> G	Yes
6308	T -> A	No
6403	G -> A	Yes
6442	G ->	No
6495	C -> T	No

Variant protein M79217_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M79217_PEA_1_T8. An alignment is given to the known protein (Exostosin-like 3) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M79217_PEA_1_P2 and EXL3_HUMAN:

1.An isolated chimeric polypeptide encoding for M79217_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC RLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYYVTENADIA CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARD RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK corresponding to amino acids 1 - 807 of EXL3_HUMAN, which also corresponds to amino acids 1 - 807 of M79217_PEA_1_P2, and a second amino acid sequence being at least 90 % homologous to AIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFHERHK CINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 820 - 919 of EXL3_HUMAN, which also corresponds to amino acids 808 - 907 of M79217_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of M79217_PEA_1_P2, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KA, having a structure as follows: a sequence starting from any of amino acid numbers 807-x to 807; and ending at any of amino acid numbers 808+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because the Signalp_hmm software predicts that this protein has a signal anchor region.

Variant protein M79217_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 583, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 583 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
104	N -> D	No
123	N -> D	No
157	I ->	No
158	R -> Q	No
204	F -> L	No
381	A ->	No
482	A ->	No
520	F -> C	No
706	L -> P	Yes
760	V -> A	No
767	R -> L	No
771	F ->	No
837	I -> V	No
875	Y -> C	No

The glycosylation sites of variant protein M79217_PEA_1_P2, as compared to the known protein Exostosin-like 3, are described in Table 584 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 584 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
290	yes	290
592	yes	592
790	yes	790
277	yes	277

Variant protein M79217_PEA_1_P2 is encoded by the following transcript(s): M79217_PEA_1_T8, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M79217_PEA_1_T8 is shown in bold; this coding portion starts at position 748 and ends at position 3468. The transcript also has the following SNPs as listed in Table 585 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 585 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
688	C -> T	No
689	T ->	No
746	T -> C	No
906	T -> A	No
1057	A -> G	No

1114	A -> G	No
1218	C ->	No
1220	G -> A	No
1359	T -> G	No
1889	C ->	No
1974	A -> G	Yes
2157	T -> C	No
2192	C ->	No
2306	T -> G	No
2864	T -> C	Yes
3026	T -> C	No
3047	G -> T	No
3060	C ->	No
3123	C -> T	Yes
3256	A -> G	No
3371	A -> G	No
3659	C ->	No
3659	C -> T	No
3724	G -> A	No
3725	G -> A	No
4054	T -> A	No
4224	G -> A	Yes
4410	C -> T	No
4748	C -> T	Yes
4857	C -> T	Yes
5075	G -> A	No
5283	G -> A	No
5381	G -> A	Yes
5525	G -> T	Yes
5781	A -> C	No

5915	G ->	No
5915	G -> C	No
5933	C -> G	Yes
5946	T -> A	No
6041	G -> A	Yes
6080	G ->	No
6133	C -> T	No

Variant protein M79217_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M79217_PEA_1_T10. An alignment is given to the known protein (Exostosin-like 3) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M79217_PEA_1_P4 and EXL3_HUMAN:

1.An isolated chimeric polypeptide encoding for M79217_PEA_1_P4, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PELRQPARLGLPECWDYRHEPRCPAQMGSHFIVQAGLKLLASSKPPKCWDY corresponding to amino acids 1 - 51 of M79217_PEA_1_P4, and a second amino acid sequence being at least 90 % homologous to RVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSY VMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFH ERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI corresponding to amino acids 759 - 919 of EXL3_HUMAN, which also corresponds to amino acids 52 - 212 of M79217_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of M79217_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PELRQPARLGLPECWDYRHEPRCPAQMGSHFIVQAGLKLLASSKPPKCWDY of M79217_PEA_1_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein M79217_PEA_1_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 586, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 586 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
53	V -> A	No
60	R -> L	No
64	F ->	No
142	I -> V	No
180	Y -> C	No

The glycosylation sites of variant protein M79217_PEA_1_P4, as compared to the known protein Exostosin-like 3, are described in Table 587 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 587 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
290	no	
592	no	
790	yes	83
277	no	

Variant protein M79217_PEA_1_P4 is encoded by the following transcript(s): M79217_PEA_1_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M79217_PEA_1_T10 is shown in bold; this coding portion starts at position 1 and ends at position 637. The transcript also has the following SNPs as listed in Table 588 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 588 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
159	T -> C	No
180	G -> T	No
193	C ->	No
256	C -> T	Yes
425	A -> G	No
540	A -> G	No
828	C ->	No
828	C -> T	No
893	G -> A	No
894	G -> A	No

1223	T -> A	No
1393	G -> A	Yes
1579	C -> T	No
1917	C -> T	Yes
2026	C -> T	Yes
2244	G -> A	No
2452	G -> A	No
2550	G -> A	Yes
2694	G -> T	Yes
2950	A -> C	No
3084	G ->	No
3084	G -> C	No
3102	C -> G	Yes
3115	T -> A	No
3210	G -> A	Yes
3249	G ->	No
3302	C -> T	No

Variant protein M79217_PEA_1_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M79217_PEA_1_T15. An alignment is given to the known protein (Exostosin-like 3) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M79217_PEA_1_P8 and EXL3_HUMAN:

1.An isolated chimeric polypeptide encoding for M79217_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEAD EAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACK KSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGC

RLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYVTENADIA
CLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTG
RAMVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESL
RSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEW
ALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPVVLGEQVQLPY
QDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTV
LAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYL
RNFTLTVTDFYRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF
QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPKLPSEDLL
WPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARD
RIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK corresponding to amino acids 1 - 807 of EXL3_HUMAN, which also corresponds to amino acids 1 - 807 of M79217_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRKSW corresponding to amino acids 808 - 812 of M79217_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M79217_PEA_1_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRKSW in M79217_PEA_1_P8.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because the Signalp_hmm software predicts that this protein has a signal anchor region.

Variant protein M79217_PEA_1_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 589, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P8

sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 589 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
104	N -> D	No
123	N -> D	No
157	I ->	No
158	R -> Q	No
204	F -> L	No
381	A ->	No
482	A ->	No
520	F -> C	No
706	L -> P	Yes
760	V -> A	No
767	R -> L	No
771	F ->	No

The glycosylation sites of variant protein M79217_PEA_1_P8, as compared to the known protein Exostosin-like 3, are described in Table 590 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 590 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
290	yes	290
592	yes	592
790	yes	790

277	yes	277

Variant protein M79217_PEA_1_P8 is encoded by the following transcript(s): M79217_PEA_1_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M79217_PEA_1_T15 is shown in bold; this coding portion starts at position 748 and ends at position 3183. The transcript also has the following SNPs as listed in Table 591 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 591 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
688	C -> T	No
689	T ->	No
746	T -> C	No
906	T -> A	No
1057	A -> G	No
1114	A -> G	No
1218	C ->	No
1220	G -> A	No
1359	T -> G	No
1889	C ->	No
1974	A -> G	Yes
2157	T -> C	No
2192	C ->	No
2306	T -> G	No
2864	T -> C	Yes
3026	T -> C	No
3047	G -> T	No

3060	C ->	No
3123	C -> T	Yes
3391	C -> T	No
3560	T -> C	No

Variant protein M79217_PEA_1_P11 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M79217_PEA_1_T18. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Signal peptide,NN:NO) predicts that this protein has a signal peptide.

Variant protein M79217_PEA_1_P11 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 592, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 592 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
17	P ->	No
28	C -> S	No
72	V ->	No
90	S -> F	No

Variant protein M79217_PEA_1_P11 is encoded by the following transcript(s): M79217_PEA_1_T18, for which the sequence(s) is/are given at the end of the application. The

coding portion of transcript M79217_PEA_1_T18 is shown in bold; this coding portion starts at position 1354 and ends at position 1674. The transcript also has the following SNPs as listed in Table 593 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M79217_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 593 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
688	C -> T	No
689	T ->	No
746	T -> C	No
772	G -> A	No
870	G -> A	Yes
1014	G -> T	Yes
1270	A -> C	No
1404	G ->	No
1404	G -> C	No
1422	C -> G	Yes
1435	T -> A	No
1530	G -> A	Yes
1569	G ->	No
1622	C -> T	No

As noted above, cluster M79217 features 32 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are

of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster M79217_PEA_1_node_2 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T3. Table 594 below describes the starting and ending position of this segment on each transcript.

Table 594 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T3	50	177

Segment cluster M79217_PEA_1_node_4 according to the present invention is supported by 8 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T8, M79217_PEA_1_T15 and M79217_PEA_1_T18. Table 595 below describes the starting and ending position of this segment on each transcript.

Table 595 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T8	1	177
M79217_PEA_1_T15	1	177
M79217_PEA_1_T18	1	177

Segment cluster M79217_PEA_1_node_9 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1. Table 596 below describes the starting and ending position of this segment on each transcript.

Table 596 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	1	597

Segment cluster M79217_PEA_1_node_10 according to the present invention is supported by 33 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T15 and M79217_PEA_1_T18. Table 597 below describes the starting and ending position of this segment on each transcript.

Table 597- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	598	1080
M79217_PEA_1_T3	272	754
M79217_PEA_1_T8	272	754
M79217_PEA_1_T15	272	754
M79217_PEA_1_T18	272	754

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 598.

Table 598 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
M79217_0_9_0	lung malignant tumors	LUN

Segment cluster M79217_PEA_1_node_11 according to the present invention is supported by 42 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T15. Table 599 below describes the starting and ending position of this segment on each transcript.

Table 599 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	1081	1523
M79217_PEA_1_T3	755	1197
M79217_PEA_1_T8	755	1197
M79217_PEA_1_T15	755	1197

Segment cluster M79217_PEA_1_node_13 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T15. Table 600 below describes the starting and ending position of this segment on each transcript.

Table 600 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	1548	2075
M79217_PEA_1_T3	1222	1749
M79217_PEA_1_T8	1222	1749
M79217_PEA_1_T15	1222	1749

Segment cluster M79217_PEA_1_node_14 according to the present invention is supported by 65 libraries. The number of libraries was determined as previously described. This

segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T15. Table 601 below describes the starting and ending position of this segment on each transcript.

Table 601 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	2076	3221
M79217_PEA_1_T3	1750	2895
M79217_PEA_1_T8	1750	2895
M79217_PEA_1_T15	1750	2895

Segment cluster M79217_PEA_1_node_16 according to the present invention is supported by 51 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T15. Table 602 below describes the starting and ending position of this segment on each transcript.

Table 602 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3222	3349
M79217_PEA_1_T3	2896	3023
M79217_PEA_1_T8	2896	3023
M79217_PEA_1_T15	2896	3023

Segment cluster M79217_PEA_1_node_23 according to the present invention is supported by 50 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3,

M79217_PEA_1_T8, M79217_PEA_1_T10 and M79217_PEA_1_T15. Table 603 below describes the starting and ending position of this segment on each transcript.

Table 603 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3350	3494
M79217_PEA_1_T3	3024	3168
M79217_PEA_1_T8	3024	3168
M79217_PEA_1_T10	157	301
M79217_PEA_1_T15	3024	3168

Segment cluster M79217_PEA_1_node_24 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T15. Table 604 below describes the starting and ending position of this segment on each transcript.

Table 604 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T15	3169	3580

Segment cluster M79217_PEA_1_node_31 according to the present invention is supported by 50 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 605 below describes the starting and ending position of this segment on each transcript.

Table 605 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3716	3960
M79217_PEA_1_T3	3390	3634
M79217_PEA_1_T8	3354	3598
M79217_PEA_1_T10	523	767

Segment cluster M79217_PEA_1_node_33 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 606 below describes the starting and ending position of this segment on each transcript.

Table 606 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	4015	4631
M79217_PEA_1_T3	3689	4305
M79217_PEA_1_T8	3653	4269
M79217_PEA_1_T10	822	1438

Segment cluster M79217_PEA_1_node_34 according to the present invention is supported by 51 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 607 below describes the starting and ending position of this segment on each transcript.

Table 607 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	4632	4869
M79217_PEA_1_T3	4306	4543
M79217_PEA_1_T8	4270	4507
M79217_PEA_1_T10	1439	1676

Segment cluster M79217_PEA_1_node_35 according to the present invention is supported by 53 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 608 below describes the starting and ending position of this segment on each transcript.

Table 608 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	4870	4997
M79217_PEA_1_T3	4544	4671
M79217_PEA_1_T8	4508	4635
M79217_PEA_1_T10	1677	1804

Segment cluster M79217_PEA_1_node_37 according to the present invention is supported by 58 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 609 below describes the starting and ending position of this segment on each transcript.

Table 609 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	5039	5280
M79217_PEA_1_T3	4713	4954
M79217_PEA_1_T8	4677	4918
M79217_PEA_1_T10	1846	2087

Segment cluster M79217_PEA_1_node_38 according to the present invention is supported by 62 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 610 below describes the starting and ending position of this segment on each transcript.

Table 610 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	5281	5436
M79217_PEA_1_T3	4955	5110
M79217_PEA_1_T8	4919	5074
M79217_PEA_1_T10	2088	2243

Segment cluster M79217_PEA_1_node_41 according to the present invention is supported by 171 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T10 and M79217_PEA_1_T18. Table 611 below describes the starting and ending position of this segment on each transcript.

Table 611- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	5628	6357
M79217_PEA_1_T3	5302	6031
M79217_PEA_1_T8	5266	5995
M79217_PEA_1_T10	2435	3164
M79217_PEA_1_T18	755	1484

Segment cluster M79217_PEA_1_node_44 according to the present invention is supported by 89 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T10 and M79217_PEA_1_T18. Table 612 below describes the starting and ending position of this segment on each transcript.

Table 612 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	6472	6659
M79217_PEA_1_T3	6146	6333
M79217_PEA_1_T8	6110	6297
M79217_PEA_1_T10	3279	3466
M79217_PEA_1_T18	1599	1786

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster M79217_PEA_1_node_0 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T3. Table 613 below describes the starting and ending position of this segment on each transcript.

Table 613 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T3	1	49

Segment cluster M79217_PEA_1_node_7 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T15 and M79217_PEA_1_T18. Table 614 below describes the starting and ending position of this segment on each transcript.

Table 614 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T3	178	271
M79217_PEA_1_T8	178	271
M79217_PEA_1_T15	178	271
M79217_PEA_1_T18	178	271

Segment cluster M79217_PEA_1_node_12 according to the present invention can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T15. Table 615 below describes the starting and ending position of this segment on each transcript.

Table 615 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	1524	1547
M79217_PEA_1_T3	1198	1221

M79217_PEA_1_T8	1198	1221
M79217_PEA_1_T15	1198	1221

Segment cluster M79217_PEA_1_node_19 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T10. Table 616 below describes the starting and ending position of this segment on each transcript.

Table 616 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T10	1	79

Segment cluster M79217_PEA_1_node_21 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T10. Table 617 below describes the starting and ending position of this segment on each transcript.

Table 617 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T10	80	156

Segment cluster M79217_PEA_1_node_26 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3 and M79217_PEA_1_T10. Table 618 below describes the starting and ending position of this segment on each transcript.

Table 618 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3495	3530
M79217_PEA_1_T3	3169	3204
M79217_PEA_1_T10	302	337

Segment cluster M79217_PEA_1_node_27 according to the present invention is supported by 46 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 619 below describes the starting and ending position of this segment on each transcript.

Table 619 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3531	3623
M79217_PEA_1_T3	3205	3297
M79217_PEA_1_T8	3169	3261
M79217_PEA_1_T10	338	430

Segment cluster M79217_PEA_1_node_30 according to the present invention is supported by 47 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 620 below describes the starting and ending position of this segment on each transcript.

Table 620 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3624	3715
M79217_PEA_1_T3	3298	3389
M79217_PEA_1_T8	3262	3353
M79217_PEA_1_T10	431	522

Segment cluster M79217_PEA_1_node_32 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 621 below describes the starting and ending position of this segment on each transcript.

Table 621 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	3961	4014
M79217_PEA_1_T3	3635	3688
M79217_PEA_1_T8	3599	3652
M79217_PEA_1_T10	768	821

Segment cluster M79217_PEA_1_node_36 according to the present invention is supported by 42 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 622 below describes the starting and ending position of this segment on each transcript.

Table 622 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	4998	5038
M79217_PEA_1_T3	4672	4712
M79217_PEA_1_T8	4636	4676
M79217_PEA_1_T10	1805	1845

Segment cluster M79217_PEA_1_node_39 according to the present invention is supported by 57 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 623 below describes the starting and ending position of this segment on each transcript.

Table 623 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	5437	5520
M79217_PEA_1_T3	5111	5194
M79217_PEA_1_T8	5075	5158
M79217_PEA_1_T10	2244	2327

Segment cluster M79217_PEA_1_node_40 according to the present invention is supported by 59 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8 and M79217_PEA_1_T10. Table 624 below describes the starting and ending position of this segment on each transcript.

Table 624 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	5521	5627
M79217_PEA_1_T3	5195	5301
M79217_PEA_1_T8	5159	5265
M79217_PEA_1_T10	2328	2434

Segment cluster M79217_PEA_1_node_42 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T10 and M79217_PEA_1_T18. Table 625 below describes the starting and ending position of this segment on each transcript.

Table 625 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	6358	6443
M79217_PEA_1_T3	6032	6117
M79217_PEA_1_T8	5996	6081
M79217_PEA_1_T10	3165	3250
M79217_PEA_1_T18	1485	1570

Segment cluster M79217_PEA_1_node_43 according to the present invention is supported by 90 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M79217_PEA_1_T1, M79217_PEA_1_T3, M79217_PEA_1_T8, M79217_PEA_1_T10 and M79217_PEA_1_T18. Table 626 below describes the starting and ending position of this segment on each transcript.

Table 626 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M79217_PEA_1_T1	6444	6471
M79217_PEA_1_T3	6118	6145
M79217_PEA_1_T8	6082	6109
M79217_PEA_1_T10	3251	3278
M79217_PEA_1_T18	1571	1598

Variant protein alignment to the previously known protein:

Sequence name: BAA25445

Sequence documentation:

Alignment of: M79217_PEA_1_P1 x BAA25445 ..

Alignment segment 1/1:

Quality: 9101.00

Escore: 0

Matching length: 919 Total length: 919

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

```
  1 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 13 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 62
            .         .         .         .         .
 51 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 63 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 112
            .         .         .         .         .
101 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
113 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 162
            .         .         .         .         .
151 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
163 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 212
            .         .         .         .         .
201 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
213 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 262
            .         .         .         .         .
251 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
263 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 312
            .         .         .         .         .
301 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
313 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 362
            .         .         .         .         .
351 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
363 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 412
```

```
                .         .         .         .         .
401 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 450
    |||||||||||||||||||||||||||||||||||||||||||||||||||
413 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 462
                .         .         .         .         .
451 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 500
    |||||||||||||||||||||||||||||||||||||||||||||||||||
463 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 512
                .         .         .         .         .
501 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 550
    |||||||||||||||||||||||||||||||||||||||||||||||||||
513 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 562
                .         .         .         .         .
551 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 600
    |||||||||||||||||||||||||||||||||||||||||||||||||||
563 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 612
                .         .         .         .         .
601 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 650
    |||||||||||||||||||||||||||||||||||||||||||||||||||
613 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 662
                .         .         .         .         .
651 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 700
    |||||||||||||||||||||||||||||||||||||||||||||||||||
663 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 712
                .         .         .         .         .
701 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 750
    |||||||||||||||||||||||||||||||||||||||||||||||||||
713 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 762
                .         .         .         .         .
751 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 800
    |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
763 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 812
                .         .         .         .         .
801 GAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIK 850
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
813 GAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIK 862
                .         .         .         .         .
851 VTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVD 900
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
863 VTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVD 912
                .
901 SVLFKTRLPHDKTKCFKFI                                 919
    |||||||||||||||||||
913 SVLFKTRLPHDKTKCFKFI                                 931
```

Sequence name: EXL3_HUMAN

Sequence documentation:

Alignment of: M79217_PEA_1_P2 x EXL3_HUMAN ..

Alignment segment 1/1:

Quality: 8873.00

Escore: 0

Matching length: 907 Total length: 919

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 98.69 Total Percent Identity: 98.69

Gaps: 1

Alignment:

```
                .         .         .         .         .
  1 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 50
                .         .         .         .         .
 51 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 100
                .         .         .         .         .
101 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 150
                .         .         .         .         .
151 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 200
                .         .         .         .         .
201 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 250
                .         .         .         .         .
251 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 300
```

```
                .         .         .         .         .
301 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 350
                .         .         .         .         .
351 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 400
                .         .         .         .         .
401 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 450
                .         .         .         .         .
451 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 500
                .         .         .         .         .
501 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 550
                .         .         .         .         .
551 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 600
                .         .         .         .         .
601 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
601 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 650
                .         .         .         .         .
651 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
651 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 700
                .         .         .         .         .
701 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 750
    |||||||||||||||||||||||||||||||||||||||||||||||||||
701 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 750
                .         .         .         .         .
751 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 800
    |||||||||||||||||||||||||||||||||||||||||||||||||||
751 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 800
                .         .         .         .         .
801 GAAFFHK............AIRDMVDEYINCEDIAMNFLVSHITRKPPIK 838
    |||||||             |||||||||||||||||||||||||||||||
801 GAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIK 850
                .         .         .         .         .
839 VTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVD 888
    |||||||||||||||||||||||||||||||||||||||||||||||||||
851 VTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVD 900
                .
889 SVLFKTRLPHDKTKCFKFI                                907
    |||||||||||||||||||
901 SVLFKTRLPHDKTKCFKFI                                919
```

Sequence name: EXL3_HUMAN

Sequence documentation:

Alignment of: M79217_PEA_1_P4 x EXL3_HUMAN ..

Alignment segment 1/1:

Quality: 1668.00

Escore: 0

Matching length: 162 Total length: 162

Matching Percent Similarity: 100.00 Matching Percent Identity: 99.38

Total Percent Similarity: 100.00 Total Percent Identity: 99.38

Gaps: 0

Alignment:

```
                 .         .         .         .         .
  51 YRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK 100
     :||||||||||||||||||||||||||||||||||||||||||||||||||
 758 FRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHK 807
                 .         .         .         .         .
 101 YYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTF 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 808 YYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTF 857
                 .         .         .         .         .
 151 RCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTR 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 858 RCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTR 907
                 .
 201 LPHDKTKCFKFI                                       212
     ||||||||||||
 908 LPHDKTKCFKFI                                       919
```

Sequence name: EXL3_HUMAN

Sequence documentation:

Alignment of: M79217_PEA_1_P8 x EXL3_HUMAN ..

Alignment segment 1/1:

Quality: 7947.00
Escore: 0
Matching length: 807 Total length: 807
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYY 50
                .         .         .         .         .
       51 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 100
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
          51 LTTLDEADEAGKRIFGPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKR 100
                         .         .         .         .         .
         101 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         101 QELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQ 150
                         .         .         .         .         .
         151 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         151 PKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYD 200
                         .         .         .         .         .
         201 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         201 SDQFVFGSYLDPLVKQAFQATARANVYVTENADIACLYVILVGEMQEPVV 250
                         .         .         .         .         .
         251 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         251 LRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQ 300
                         .         .         .         .         .
         301 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 350
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         301 STFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKI 350
                         .         .         .         .         .
         351 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         351 ESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKAVQDSKLDQVLVEFTC 400
                         .         .         .         .         .
         401 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 450
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
         401 KNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR 450
                         .         .         .         .         .
```

```
451 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSL 500
              .         .         .         .         .
501 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 SDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQIPAAPIREEAA 550
              .         .         .         .         .
551 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 AEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDF 600
              .         .         .         .         .
601 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 YRSWNCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEF 650
              .         .         .         .         .
651 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 QAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVVVVWNSPK 700
              .         .         .         .         .
701 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 750
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 LPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLR 750
              .         .         .         .         .
751 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 800
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 HDEIMFGFRVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLT 800

801 GAAFFHK                                             807
    |||||||
801 GAAFFHK                                             807
```

DESCRIPTION FOR CLUSTER M62096

Cluster M62096 features 9 transcript(s) and 42 segment(s) of interest, the names for which are given in Tables 627 and 628, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 629.

Table 627 - Transcripts of interest

Transcript Name	Sequence ID No.
M62096_PEA_1_T4	65
M62096_PEA_1_T5	66
M62096_PEA_1_T6	67
M62096_PEA_1_T7	68
M62096_PEA_1_T9	69
M62096_PEA_1_T11	70
M62096_PEA_1_T13	71
M62096_PEA_1_T14	72
M62096_PEA_1_T15	73

Table 628 - Segments of interest

Segment Name	Sequence ID No.
M62096_PEA_1_node_0	616
M62096_PEA_1_node_2	617
M62096_PEA_1_node_15	618
M62096_PEA_1_node_17	619
M62096_PEA_1_node_19	620
M62096_PEA_1_node_23	621
M62096_PEA_1_node_27	623
M62096_PEA_1_node_29	624
M62096_PEA_1_node_31	625
M62096_PEA_1_node_34	626

M62096_PEA_1_node_36	627
M62096_PEA_1_node_38	628
M62096_PEA_1_node_40	629
M62096_PEA_1_node_48	630
M62096_PEA_1_node_50	631
M62096_PEA_1_node_56	632
M62096_PEA_1_node_60	633
M62096_PEA_1_node_65	634
M62096_PEA_1_node_69	635
M62096_PEA_1_node_71	636
M62096_PEA_1_node_1	637
M62096_PEA_1_node_4	638
M62096_PEA_1_node_6	639
M62096_PEA_1_node_7	640
M62096_PEA_1_node_9	641
M62096_PEA_1_node_11	642
M62096_PEA_1_node_13	643
M62096_PEA_1_node_21	644
M62096_PEA_1_node_25	645
M62096_PEA_1_node_33	646
M62096_PEA_1_node_42	647
M62096_PEA_1_node_44	648
M62096_PEA_1_node_47	649
M62096_PEA_1_node_51	650
M62096_PEA_1_node_53	651
M62096_PEA_1_node_55	652
M62096_PEA_1_node_58	653
M62096_PEA_1_node_62	654
M62096_PEA_1_node_66	655
M62096_PEA_1_node_67	656

M62096_PEA_1_node_68	657
M62096_PEA_1_node_70	658

Table 629 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
M62096_PEA_1_P4	1341	M62096_PEA_1_T6
M62096_PEA_1_P5	1342	M62096_PEA_1_T7
M62096_PEA_1_P3	1343	M62096_PEA_1_T9
M62096_PEA_1_P7	1344	M62096_PEA_1_T11
M62096_PEA_1_P8	1345	M62096_PEA_1_T13
M62096_PEA_1_P9	1346	M62096_PEA_1_T14
M62096_PEA_1_P10	1347	M62096_PEA_1_T15
M62096_PEA_1_P11	1348	M62096_PEA_1_T4
M62096_PEA_1_P12	1349	M62096_PEA_1_T5

These sequences are variants of the known protein Kinesin heavy chain isoform 5C (SwissProt accession identifier KF5C_HUMAN; known also according to the synonyms Kinesin heavy chain neuron-specific 2), SEQ ID NO: 1438, referred to herein as the previously known protein.

Protein Kinesin heavy chain isoform 5C is known or believed to have the following function(s): Kinesin is a microtubule-associated force-producing protein that may play a role in organelle transport. The sequence for protein Kinesin heavy chain isoform 5C is given at the end of the application, as "Kinesin heavy chain isoform 5C amino acid sequence". Known polymorphisms for this sequence are as shown in Table 630.

Table 630 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
355 - 360	TLKNVI -> STHASV
583 - 585	EFT -> DRV

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: organelle organization and biogenesis, which are annotation(s) related to Biological Process; microtubule motor; ATP binding, which are annotation(s) related to Molecular Function; and kinesin, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster M62096 features 9 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Kinesin heavy chain isoform 5C. A description of each variant protein according to the present invention is now provided.

Variant protein M62096_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T6. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P4 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P4, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MATYIH corresponding to amino acids 1 - 6 of M62096_PEA_1_P4, and a second amino acid sequence being at least 90 % homologous to VSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNC RTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKT LKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKE KYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQ IENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRE

LSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYIS
KMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQN
MEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQ
MESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREM
KLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDD
GGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALES
ALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTA
VHAIRGGGGSSSNSTHYQK corresponding to amino acids 239 - 957 of KF5C_HUMAN, which also corresponds to amino acids 7 - 725 of M62096_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of M62096_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MATYIH of M62096_PEA_1_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P4 is encoded by the following transcript(s): M62096_PEA_1_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T6 is shown in bold; this coding portion starts at position 108 and ends at position 2282. The transcript also has the following SNPs as listed in Table 631 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 631 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
5757	G -> T	No

Variant protein M62096_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T7. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P5 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MTRILQDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWK KKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNI APVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRR DYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDEL AQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNG VIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHE AKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQ DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDY NKLKIEDQEREMKLEKILLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTT RVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRL RATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPI RPGHYPASSPTAVHAIRGGGGSSSNSTHYQK corresponding to amino acids 284 - 957 of KF5C_HUMAN, which also corresponds to amino acids 1 - 674 of M62096_PEA_1_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P5 is encoded by the following transcript(s): M62096_PEA_1_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T7 is shown in bold; this coding portion starts at position 283 and ends at position 2304. The transcript also has the following SNPs as listed in Table 632 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 632 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
5779	G -> T	No

Variant protein M62096_PEA_1_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T9. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P3 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P3, comprising a first amino acid sequence being at least 90 % homologous to MELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSL YRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKD EVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQELS NHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKS LVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQL EESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAH QKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLN DKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQK QKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKEN AMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGG SSSNSTHYQK corresponding to amino acids 365 - 957 of KF5C_HUMAN, which also corresponds to amino acids 1 - 593 of M62096_PEA_1_P3.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P3 is encoded by the following transcript(s): M62096_PEA_1_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T9 is shown in bold; this coding portion starts at position 565 and ends at position 2343. The transcript also has the following SNPs as listed in Table 633 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 633 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
5818	G -> T	No

Variant protein M62096_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T11. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P7 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P7, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MTQNFRLMWNILLFPLNFS corresponding to amino acids 1 - 19 of M62096_PEA_1_P7, and a second amino acid sequence being at least 90 % homologous to LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL QTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVR DNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRA KNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK corresponding to amino acids 738 - 957 of KF5C_HUMAN, which also corresponds to amino acids 20 - 239 of M62096_PEA_1_P7, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of M62096_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MTQNFRLMWNILLFPLNFS of M62096_PEA_1_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-peptide prediction programs (HMM:Non-secretory protein,NN:YES) predicts that this protein has a signal peptide.

Variant protein M62096_PEA_1_P7 is encoded by the following transcript(s): M62096_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T11 is shown in bold; this coding portion starts at position 633 and ends at position 1349. The transcript also has the following SNPs as listed in Table 634 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 634 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
4824	G -> T	No

Variant protein M62096_PEA_1_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T13. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P8 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK TLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRL QIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQR ELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYI SKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQN MEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQ MESHREAHQKQLSRLRDEIEEKQKIIDEIR corresponding to amino acids 1 - 736 of KF5C_HUMAN, which also corresponds to amino acids 1 - 736 of M62096_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence E corresponding to amino acids 737 - 737 of M62096_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 635, (given according to their position(s) on the

amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 635 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> T	Yes

Variant protein M62096_PEA_1_P8 is encoded by the following transcript(s): M62096_PEA_1_T13, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T13 is shown in bold; this coding portion starts at position 396 and ends at position 2606. The transcript also has the following SNPs as listed in Table 636 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 636 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
92	C -> A	Yes
408	G -> A	Yes

Variant protein M62096_PEA_1_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T14. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P9 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P9, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK TLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDE corresponding to amino acids 1 - 454 of KF5C_HUMAN, which also corresponds to amino acids 1 - 454 of M62096_PEA_1_P9, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VKNAIYFFFHKVLLLLFVVDVCSRNLIGIEAFHNYRIMWKFLGRCPFTASYKLIITEFRK corresponding to amino acids 455 - 514 of M62096_PEA_1_P9, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M62096_PEA_1_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VKNAIYFFFHKVLLLLFVVDVCSRNLIGIEAFHNYRIMWKFLGRCPFTASYKLIITEFRK in M62096_PEA_1_P9.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P9 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 637, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 637 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> T	Yes

Variant protein M62096_PEA_1_P9 is encoded by the following transcript(s): M62096_PEA_1_T14, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T14 is shown in bold; this coding portion starts at position 396 and ends at position 1937. The transcript also has the following SNPs as listed in Table 638 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 638 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
92	C -> A	Yes
408	G -> A	Yes

Variant protein M62096_PEA_1_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T15. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously

published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P10 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P10, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MTQNFRLMWNILLFPLNFS corresponding to amino acids 1 - 19 of M62096_PEA_1_P10, a second amino acid sequence being at least 90 % homologous to LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL QTLHNLRKLFVQDLTTRVKK corresponding to amino acids 738 - 815 of KF5C_HUMAN, which also corresponds to amino acids 20 - 97 of M62096_PEA_1_P10, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSSLCLNGTEKKIKDGREESFSVEISLA corresponding to amino acids 98 - 125 of M62096_PEA_1_P10, wherein said first amino acid sequence, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of M62096_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MTQNFRLMWNILLFPLNFS of M62096_PEA_1_P10.

3.An isolated polypeptide encoding for a tail of M62096_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSSLCLNGTEKKIKDGREESFSVEISLA in M62096_PEA_1_P10.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because one of the two signal-

peptide prediction programs (HMM:Non-secretory protein,NN:YES) predicts that this protein has a signal peptide.

Variant protein M62096_PEA_1_P10 is encoded by the following transcript(s): M62096_PEA_1_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T15 is shown in bold; this coding portion starts at position 633 and ends at position 1007.

Variant protein M62096_PEA_1_P11 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T4. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P11 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNK TLKNVIQHLEMELNRWRN corresponding to amino acids 1 - 372 of KF5C_HUMAN, which also corresponds to amino acids 1 - 372 of M62096_PEA_1_P11, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DFLAAHVFGKLLE corresponding to amino acids 373 - 385 of M62096_PEA_1_P11, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M62096_PEA_1_P11, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DFLAAHVFGKLLE in M62096_PEA_1_P11.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P11 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 639, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 639 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> T	Yes

Variant protein M62096_PEA_1_P11 is encoded by the following transcript(s): M62096_PEA_1_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T4 is shown in bold; this coding portion starts at position 396 and ends at position 1550. The transcript also has the following SNPs as listed in Table 640 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein M62096_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 640 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
92	C -> A	Yes
408	G -> A	Yes
6908	G -> T	No

Variant protein M62096_PEA_1_P12 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M62096_PEA_1_T5. An alignment is given to the known protein (Kinesin heavy chain isoform 5C) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M62096_PEA_1_P12 and KF5C_HUMAN:

1.An isolated chimeric polypeptide encoding for M62096_PEA_1_P12, comprising a first amino acid sequence being at least 90 % homologous to MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQ EQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFD HIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNLAVHEDKNRVPYVKGCTERFVSSPE EVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSE KVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGN CRTTIVICCSPSVFNEAETKSTLMFGQR corresponding to amino acids 1 - 323 of KF5C_HUMAN, which also corresponds to amino acids 1 - 323 of M62096_PEA_1_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence V corresponding to amino acids 324 - 324 of M62096_PEA_1_P12,

wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein M62096_PEA_1_P12 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 641, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 641 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> T	Yes

Variant protein M62096_PEA_1_P12 is encoded by the following transcript(s): M62096_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M62096_PEA_1_T5 is shown in bold; this coding portion starts at position 378 and ends at position 1349. The transcript also has the following SNPs as listed in Table 642 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M62096_PEA_1_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 642 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
92	C -> A	Yes
390	G -> A	Yes
6784	G -> T	No

As noted above, cluster M62096 features 42 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster M62096_PEA_1_node_0 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 643 below describes the starting and ending position of this segment on each transcript.

Table 643 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1	355
M62096_PEA_1_T5	1	355
M62096_PEA_1_T13	1	355
M62096_PEA_1_T14	1	355

Segment cluster M62096_PEA_1_node_2 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 644 below describes the starting and ending position of this segment on each transcript.

Table 644 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	374	521
M62096_PEA_1_T5	356	503
M62096_PEA_1_T13	374	521
M62096_PEA_1_T14	374	521

Segment cluster M62096_PEA_1_node_15 according to the present invention is supported by 28 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 645 below describes the starting and ending position of this segment on each transcript.

Table 645 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	985	1109
M62096_PEA_1_T5	967	1091
M62096_PEA_1_T13	985	1109
M62096_PEA_1_T14	985	1109

Segment cluster M62096_PEA_1_node_17 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T7. Table 646 below describes the starting and ending position of this segment on each transcript.

Table 646 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T7	1	147

Segment cluster M62096_PEA_1_node_19 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T6 and M62096_PEA_1_T9. Table 647 below describes the starting and ending position of this segment on each transcript.

Table 647 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T6	1	125
M62096_PEA_1_T9	1	125

Segment cluster M62096_PEA_1_node_23 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 648 below describes the starting and ending position of this segment on each transcript.

Table 648 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1215	1363
M62096_PEA_1_T5	1197	1345
M62096_PEA_1_T6	231	379

M62096_PEA_1_T7	253	401
M62096_PEA_1_T9	231	379
M62096_PEA_1_T13	1215	1363
M62096_PEA_1_T14	1215	1363

Segment cluster M62096_PEA_1_node_27 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 649 below describes the starting and ending position of this segment on each transcript.

Table 649 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1364	1512
M62096_PEA_1_T5	1407	1555
M62096_PEA_1_T6	380	528
M62096_PEA_1_T7	402	550
M62096_PEA_1_T9	441	589
M62096_PEA_1_T13	1364	1512
M62096_PEA_1_T14	1364	1512

Segment cluster M62096_PEA_1_node_29 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4. Table 650 below describes the starting and ending position of this segment on each transcript.

Table 650 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1513	1679

Segment cluster M62096_PEA_1_node_31 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 651 below describes the starting and ending position of this segment on each transcript.

Table 651 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1680	1855
M62096_PEA_1_T5	1556	1731
M62096_PEA_1_T6	529	704
M62096_PEA_1_T7	551	726
M62096_PEA_1_T9	590	765
M62096_PEA_1_T13	1513	1688
M62096_PEA_1_T14	1513	1688

Segment cluster M62096_PEA_1_node_34 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T14. Table 652 below describes the starting and ending position of this segment on each transcript.

Table 652 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T14	1758	2261

Segment cluster M62096_PEA_1_node_36 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 653 below describes the starting and ending position of this segment on each transcript.

Table 653 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1925	2131
M62096_PEA_1_T5	1801	2007
M62096_PEA_1_T6	774	980
M62096_PEA_1_T7	796	1002
M62096_PEA_1_T9	835	1041
M62096_PEA_1_T13	1758	1964

Segment cluster M62096_PEA_1_node_38 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 654 below describes the starting and ending position of this segment on each transcript.

Table 654 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

M62096_PEA_1_T4	2132	2278
M62096_PEA_1_T5	2008	2154
M62096_PEA_1_T6	981	1127
M62096_PEA_1_T7	1003	1149
M62096_PEA_1_T9	1042	1188
M62096_PEA_1_T13	1965	2111

Segment cluster M62096_PEA_1_node_40 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 655 below describes the starting and ending position of this segment on each transcript.

Table 655 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2279	2467
M62096_PEA_1_T5	2155	2343
M62096_PEA_1_T6	1128	1316
M62096_PEA_1_T7	1150	1338
M62096_PEA_1_T9	1189	1377
M62096_PEA_1_T13	2112	2300

Segment cluster M62096_PEA_1_node_48 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T13. Table 656 below describes the starting and ending position of this segment on each transcript.

Table 656 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T13	2606	2945

Segment cluster M62096_PEA_1_node_50 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T11 and M62096_PEA_1_T15. Table 657 below describes the starting and ending position of this segment on each transcript.

Table 657 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T11	1	688
M62096_PEA_1_T15	1	688

Segment cluster M62096_PEA_1_node_56 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T15. Table 658 below describes the starting and ending position of this segment on each transcript.

Table 658 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T15	924	1059

Segment cluster M62096_PEA_1_node_60 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This

segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 659 below describes the starting and ending position of this segment on each transcript.

Table 659 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	3113	3329
M62096_PEA_1_T5	2989	3205
M62096_PEA_1_T6	1962	2178
M62096_PEA_1_T7	1984	2200
M62096_PEA_1_T9	2023	2239
M62096_PEA_1_T11	1029	1245

Segment cluster M62096_PEA_1_node_65 according to the present invention is supported by 51 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 660 below describes the starting and ending position of this segment on each transcript.

Table 660 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	3444	4763
M62096_PEA_1_T5	3320	4639
M62096_PEA_1_T6	2293	3612
M62096_PEA_1_T7	2315	3634
M62096_PEA_1_T9	2354	3673
M62096_PEA_1_T11	1360	2679

Segment cluster M62096_PEA_1_node_69 according to the present invention is supported by 85 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 661 below describes the starting and ending position of this segment on each transcript.

Table 661 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	4894	5826
M62096_PEA_1_T5	4770	5702
M62096_PEA_1_T6	3743	4675
M62096_PEA_1_T7	3765	4697
M62096_PEA_1_T9	3804	4736
M62096_PEA_1_T11	2810	3742

Segment cluster M62096_PEA_1_node_71 according to the present invention is supported by 178 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 662 below describes the starting and ending position of this segment on each transcript.

Table 662 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	5882	7128
M62096_PEA_1_T5	5758	7004
M62096_PEA_1_T6	4731	5977

M62096_PEA_1_T7	4753	5999
M62096_PEA_1_T9	4792	6038
M62096_PEA_1_T11	3798	5044

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster M62096_PEA_1_node_1 according to the present invention can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 663 below describes the starting and ending position of this segment on each transcript.

Table 663 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	356	373
M62096_PEA_1_T13	356	373
M62096_PEA_1_T14	356	373

Segment cluster M62096_PEA_1_node_4 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 664 below describes the starting and ending position of this segment on each transcript.

Table 664 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	522	612
M62096_PEA_1_T5	504	594

M62096_PEA_1_T13	522	612
M62096_PEA_1_T14	522	612

Segment cluster M62096_PEA_1_node_6 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 665 below describes the starting and ending position of this segment on each transcript.

Table 665 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	613	686
M62096_PEA_1_T5	595	668
M62096_PEA_1_T13	613	686
M62096_PEA_1_T14	613	686

Segment cluster M62096_PEA_1_node_7 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 666 below describes the starting and ending position of this segment on each transcript.

Table 666 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	687	791
M62096_PEA_1_T5	669	773
M62096_PEA_1_T13	687	791

M62096_PEA_1_T14	687	791

Segment cluster M62096_PEA_1_node_9 according to the present invention is supported by 18 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 667 below describes the starting and ending position of this segment on each transcript.

Table 667 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	792	840
M62096_PEA_1_T5	774	822
M62096_PEA_1_T13	792	840
M62096_PEA_1_T14	792	840

Segment cluster M62096_PEA_1_node_11 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 668 below describes the starting and ending position of this segment on each transcript.

Table 668 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	841	896
M62096_PEA_1_T5	823	878
M62096_PEA_1_T13	841	896
M62096_PEA_1_T14	841	896

Segment cluster M62096_PEA_1_node_13 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 669 below describes the starting and ending position of this segment on each transcript.

Table 669 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	897	984
M62096_PEA_1_T5	879	966
M62096_PEA_1_T13	897	984
M62096_PEA_1_T14	897	984

Segment cluster M62096_PEA_1_node_21 according to the present invention is supported by 33 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 670 below describes the starting and ending position of this segment on each transcript.

Table 670 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1110	1214
M62096_PEA_1_T5	1092	1196
M62096_PEA_1_T6	126	230
M62096_PEA_1_T7	148	252

M62096_PEA_1_T9	126	230
M62096_PEA_1_T13	1110	1214
M62096_PEA_1_T14	1110	1214

Segment cluster M62096_PEA_1_node_25 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T5 and M62096_PEA_1_T9. Table 671 below describes the starting and ending position of this segment on each transcript.

Table 671 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T5	1346	1406
M62096_PEA_1_T9	380	440

Segment cluster M62096_PEA_1_node_33 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T13 and M62096_PEA_1_T14. Table 672 below describes the starting and ending position of this segment on each transcript.

Table 672 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	1856	1924
M62096_PEA_1_T5	1732	1800
M62096_PEA_1_T6	705	773

M62096_PEA_1_T7	727	795
M62096_PEA_1_T9	766	834
M62096_PEA_1_T13	1689	1757
M62096_PEA_1_T14	1689	1757

Segment cluster M62096_PEA_1_node_42 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 673 below describes the starting and ending position of this segment on each transcript.

Table 673 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2468	2585
M62096_PEA_1_T5	2344	2461
M62096_PEA_1_T6	1317	1434
M62096_PEA_1_T7	1339	1456
M62096_PEA_1_T9	1378	1495
M62096_PEA_1_T13	2301	2418

Segment cluster M62096_PEA_1_node_44 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 674 below describes the starting and ending position of this segment on each transcript.

Table 674 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2586	2662
M62096_PEA_1_T5	2462	2538
M62096_PEA_1_T6	1435	1511
M62096_PEA_1_T7	1457	1533
M62096_PEA_1_T9	1496	1572
M62096_PEA_1_T13	2419	2495

Segment cluster M62096_PEA_1_node_47 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T13. Table 675 below describes the starting and ending position of this segment on each transcript.

Table 675 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2663	2772
M62096_PEA_1_T5	2539	2648
M62096_PEA_1_T6	1512	1621
M62096_PEA_1_T7	1534	1643
M62096_PEA_1_T9	1573	1682
M62096_PEA_1_T13	2496	2605

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 676.

Table 676 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
M62096_0_7_0	lung malignant tumors	LUN

Segment cluster M62096_PEA_1_node_51 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T11 and M62096_PEA_1_T15. Table 677 below describes the starting and ending position of this segment on each transcript.

Table 677 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2773	2874
M62096_PEA_1_T5	2649	2750
M62096_PEA_1_T6	1622	1723
M62096_PEA_1_T7	1644	1745
M62096_PEA_1_T9	1683	1784
M62096_PEA_1_T11	689	790
M62096_PEA_1_T15	689	790

Segment cluster M62096_PEA_1_node_53 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T11 and M62096_PEA_1_T15. Table 678 below describes the starting and ending position of this segment on each transcript.

Table 678 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2875	2935
M62096_PEA_1_T5	2751	2811
M62096_PEA_1_T6	1724	1784
M62096_PEA_1_T7	1746	1806
M62096_PEA_1_T9	1785	1845
M62096_PEA_1_T11	791	851
M62096_PEA_1_T15	791	851

Segment cluster M62096_PEA_1_node_55 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9, M62096_PEA_1_T11 and M62096_PEA_1_T15. Table 679 below describes the starting and ending position of this segment on each transcript.

Table 679 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	2936	3007
M62096_PEA_1_T5	2812	2883
M62096_PEA_1_T6	1785	1856
M62096_PEA_1_T7	1807	1878
M62096_PEA_1_T9	1846	1917
M62096_PEA_1_T11	852	923
M62096_PEA_1_T15	852	923

Segment cluster M62096_PEA_1_node_58 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 680 below describes the starting and ending position of this segment on each transcript.

Table 680 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	3008	3112
M62096_PEA_1_T5	2884	2988
M62096_PEA_1_T6	1857	1961
M62096_PEA_1_T7	1879	1983
M62096_PEA_1_T9	1918	2022
M62096_PEA_1_T11	924	1028

Segment cluster M62096_PEA_1_node_62 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 681 below describes the starting and ending position of this segment on each transcript.

Table 681 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	3330	3443
M62096_PEA_1_T5	3206	3319
M62096_PEA_1_T6	2179	2292
M62096_PEA_1_T7	2201	2314
M62096_PEA_1_T9	2240	2353

M62096_PEA_1_T11	1246	1359

Segment cluster M62096_PEA_1_node_66 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 682 below describes the starting and ending position of this segment on each transcript.

Table 682 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	4764	4881
M62096_PEA_1_T5	4640	4757
M62096_PEA_1_T6	3613	3730
M62096_PEA_1_T7	3635	3752
M62096_PEA_1_T9	3674	3791
M62096_PEA_1_T11	2680	2797

Segment cluster M62096_PEA_1_node_67 according to the present invention can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 683 below describes the starting and ending position of this segment on each transcript.

Table 683 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	4882	4887
M62096_PEA_1_T5	4758	4763
M62096_PEA_1_T6	3731	3736

M62096_PEA_1_T7	3753	3758
M62096_PEA_1_T9	3792	3797
M62096_PEA_1_T11	2798	2803

Segment cluster M62096_PEA_1_node_68 according to the present invention can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 684below describes the starting and ending position of this segment on each transcript.

Table 684 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	4888	4893
M62096_PEA_1_T5	4764	4769
M62096_PEA_1_T6	3737	3742
M62096_PEA_1_T7	3759	3764
M62096_PEA_1_T9	3798	3803
M62096_PEA_1_T11	2804	2809

Segment cluster M62096_PEA_1_node_70 according to the present invention is supported by 55 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M62096_PEA_1_T4, M62096_PEA_1_T5, M62096_PEA_1_T6, M62096_PEA_1_T7, M62096_PEA_1_T9 and M62096_PEA_1_T11. Table 685 below describes the starting and ending position of this segment on each transcript.

Table 685 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M62096_PEA_1_T4	5827	5881

M62096_PEA_1_T5	5703	5757
M62096_PEA_1_T6	4676	4730
M62096_PEA_1_T7	4698	4752
M62096_PEA_1_T9	4737	4791
M62096_PEA_1_T11	3743	3797

Variant protein alignment to the previously known protein:

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P4 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 6936.00

Escore: 0

Matching length: 719 Total length: 719

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                 .         .         .         .         .
       7 VSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRIL 56
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     239 VSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRIL 288
                 .         .         .         .         .
      57 QDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELT 106
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     289 QDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELT 338
                 .         .         .         .         .
     107 AEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQK 156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     339 AEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQK 388
                 .         .         .         .         .
     157 NLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQ 206
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     389 NLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQ 438
                 .         .         .         .         .
     207 SQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVL 256
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     439 SQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVL 488
                 .         .         .         .         .
     257 QALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQE 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     489 QALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQE 538
                 .         .         .         .         .
     307 LSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMAR 356
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     539 LSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMAR 588
                 .         .         .         .         .
     357 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEA 406
```

```
589 LYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEA 638

407 KIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKE 456

639 KIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKE 688

457 HLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDL 506

689 HLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDL 738

507 NQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLE 556

739 NQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLE 788

557 ETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLE 606

789 ETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLE 838

607 NNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKE 656

839 NNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKE 888

657 NAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTA 706

889 NAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTA 938

707 VHAIRGGGGSSSNSTHYQK                                 725

939 VHAIRGGGGSSSNSTHYQK                                 957
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P5 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 6520.00

Escore: 0

Matching length: 674 Total length: 674

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MTRILQDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSV 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    284 MTRILQDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSV 333
                .         .         .         .         .
     51 NLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQIS 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
334 NLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQIS 383
                .         .         .         .         .
101 AKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDD 150
    |||||||||||||||||||||||||||||||||||||||||||||||||||
384 AKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDD 433
                .         .         .         .         .
151 EINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDE 200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
434 EINQQSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDE 483
                .         .         .         .         .
201 VKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQREL 250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
484 VKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQREL 533
                .         .         .         .         .
251 SQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEE 300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
534 SQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEE 583
                .         .         .         .         .
301 FTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLI 350
    |||||||||||||||||||||||||||||||||||||||||||||||||||
584 FTMARLYISKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLI 633
                .         .         .         .         .
351 SQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQ 400
    |||||||||||||||||||||||||||||||||||||||||||||||||||
634 SQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQ 683
                .         .         .         .         .
401 DKEKEHLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIID 450
    |||||||||||||||||||||||||||||||||||||||||||||||||||
684 DKEKEHLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIID 733
                .         .         .         .         .
451 EIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQARED 500
```

745

```
                                                                                
734 EIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQARED 783
                .         .         .         .         .
501 LKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQK 550
                                                                                
784 LKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQK 833
                .         .         .         .         .
551 ISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESAL 600
                                                                                
834 ISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESAL 883
                .         .         .         .         .
601 KEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPA 650
                                                                                
884 KEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPA 933
                .         .
651 SSPTAVHAIRGGGGSSSNSTHYQK                           674
                                                        
934 SSPTAVHAIRGGGGSSSNSTHYQK                           957
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P3 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 5726.00

Escore: 0

Matching length: 593 Total length: 593

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    365 MELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEK 414
                .         .         .         .         .
     51 EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYE 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    415 EKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDYE 464
                .         .         .         .         .
    101 KIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQ 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    465 KIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQ 514
                .         .         .         .         .
    151 LTDELAQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGI 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    515 LTDELAQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGI 564
                .         .         .         .         .
    201 IGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMD 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
565 IGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMD 614
251 SNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSL 300
615 SNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSL 664
301 SEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAH 350
665 SEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAH 714
351 QKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQERE 400
715 QKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQERE 764
401 MKLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVK 450
765 MKLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVK 814
451 KSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPK 500
815 KSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPK 864
501 LEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMA 550
865 LEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMA 914
551 RRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK         593
915 RRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK         957
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P7 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 2117.00

Escore: 0

Matching length: 220 Total length: 220

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        20 LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGL 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       738 LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGL 787
                .         .         .         .         .
        70 EETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFL 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       788 EETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFL 837
```

```
                .         .         .         .         .
       120 ENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAK 169
           |||||||||||||||||||||||||||||||||||||||||||||||||||
       838 ENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAK 887
                .         .         .         .         .
       170 ENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPT 219
           |||||||||||||||||||||||||||||||||||||||||||||||||||
       888 ENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPT 937
                .         .
       220 AVHAIRGGGGSSSNSTHYQK                               239
           ||||||||||||||||||||
       938 AVHAIRGGGGSSSNSTHYQK                               957
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P8 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 7146.00

Escore: 0

Matching length: 737 Total length: 737

750

Matching Percent Similarity: 100.00 Matching Percent Identity: 99.86

Total Percent Similarity: 100.00 Total Percent Identity: 99.86

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
                .         .         .         .         .
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
                .         .         .         .         .
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
                .         .         .         .         .
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
                .         .         .         .         .
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
                .         .         .         .         .
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
```

```
                .         .         .         .         .
301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350
                .         .         .         .         .
351 EKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIID 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 EKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIID 400
                .         .         .         .         .
401 NIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQML 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 NIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQML 450
                .         .         .         .         .
451 DQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQ 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 DQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQ 500
                .         .         .         .         .
501 KSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQELSNHQKKRATEI 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 KSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQELSNHQKKRATEI 550
                .         .         .         .         .
551 LNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKS 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 LNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKS 600
                .         .         .         .         .
601 LVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNM 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
601 LVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNM 650
                .         .         .         .         .
651 EQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMK 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

752

651 EQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMK 700

. . .

701 KALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRE 737

||||||||||||||||||||||||||||||||||:

701 KALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRD 737

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P9 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 4434.00

Escore: 0

Matching length: 454 Total length: 454

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

```
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
            .         .         .         .         .
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
            .         .         .         .         .
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
            .         .         .         .         .
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
            .         .         .         .         .
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
            .         .         .         .         .
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
            .         .         .         .         .
301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350
            .         .         .         .         .
351 EKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIID 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 EKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIID 400
```

```
                .         .         .         .         .
401 NIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQML 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 NIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQML 450

451 DQDE                                                 454
    ||||
451 DQDE                                                 454
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P10 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 747.00
Escore: 0
Matching length: 78 Total length: 78
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                    .         .         .         .         .
        20 LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGL 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       738 LNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGL 787

                    .         .
        70 EETVSRELQTLHNLRKLFVQDLTTRVKK                        97
           ||||||||||||||||||||||||||||
       788 EETVSRELQTLHNLRKLFVQDLTTRVKK                       815
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P11 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 3634.00

Escore: 0

Matching length: 372 Total length: 372

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
                .         .         .         .         .
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
                .         .         .         .         .
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
                .         .         .         .         .
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
                .         .         .         .         .
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
                .         .         .         .         .
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
                .         .         .         .         .
301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350
```

757

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||
          301 VICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEK 350

                      .          .
          351 EKNKTLKNVIQHLEMELNRWRN                               372
                ||||||||||||||||||||||
          351 EKNKTLKNVIQHLEMELNRWRN                               372
```

Sequence name: KF5C_HUMAN

Sequence documentation:

Alignment of: M62096_PEA_1_P12 x KF5C_HUMAN ..

Alignment segment 1/1:

Quality: 3145.00

Escore: 0

Matching length: 323 Total length: 323

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFD 50
                .         .         .         .         .
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 RVLPPNTTQEQVYNACAKQIVKDVLEGYNGTIFAYGQTSSGKTHTMEGKL 100
                .         .         .         .         .
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 HDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVS 150
                .         .         .         .         .
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 KTNLAVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNE 200
                .         .         .         .         .
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 HSSRSHSIFLINIKQENVETEKKLSGKLYLVDLAGSEKVSKTGAEGAVLD 250
                .         .         .         .         .
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 EAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTI 300
                .         .
301 VICCSPSVFNEAETKSTLMFGQR                             323
    |||||||||||||||||||||||
301 VICCSPSVFNEAETKSTLMFGQR                             323
```

Expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) M62069 transcripts which are detectable by amplicon as depicted in sequence name M62069 seg19 in normal and cancerous lung tissues

Expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts detectable by or according to seg19, M62069 seg19 amplicon (SEQ ID NO: 1657) and M62069 seg19F (SEQ ID NO: 1655) and M62069 seg19R (SEQ ID NO: 1656) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 65 is a histogram showing over expression of the above-indicated Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.

As is evident from Figure 65, the expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 5 fold was found in 2 out of 15 adenocarcinoma samples, and in 8 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: M62069 seg19F forward primer; and M62069 seg19R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: M62069 seg19.

Forward primer -M62069 seg19F (SEQ ID NO: 1655):
GCTGATTGTCCCCATGAAGG

Reverse primer- M62069 seg19 (SEQ ID NO: 1656): TGGCATACGGGAACTCAGTG

Amplicon (SEQ ID NO: 1657):
GCTGATTGTCCCCATGAAGGCCAGCCTTGAAGCTTGGTCAGTCTCCCTAACTGTATG
ATTGATCCCCACTTATTGCACTACATCACTGAGTTCCCGTATGC

Expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) *M62069 transcripts which are detectable by amplicon as depicted in sequence name M62069 seg29 in normal and cancerous lung tissues*

Expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts detectable by or according to seg29, M62069 seg29 amplicon (SEQ ID NO: 1660) and M62069 seg29F (SEQ ID NO: 1658) and M62069 seg29R (SEQ ID NO: 1659) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of

the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 66 is a histogram showing over expression of the above-indicated Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.

As is evident from Figure 66, the expression of Homo sapiens protein tyrosine phosphatase, receptor type, S (PTPRS) transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 5 fold was found in 2 out of 15 adenocarcinoma samples, and in 7 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: M62069 seg29F forward primer; and M62069 seg29R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: M62069 seg29.

Forward primer -M62069 seg29F: ATTGAATAATTCAGCACCTGAGGC

Reverse primer- M62069 seg29R: TTCATATGGCTACTCCCCACCT

Amplicon:

ATTGAATAATTCAGCACCTGAGGCTGGTGGATGATTCTTTGCAATTTGGCAGGAATG

GGAGAGTCGGGAGCAGTAGTTGGCAAGGTGGGGAGTAGCCATATGAA

DESCRIPTION FOR CLUSTER M78076

Cluster M78076 features 9 transcript(s) and 35 segment(s) of interest, the names for which are given in Tables 686 and 687, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 688.

Table 686 - Transcripts of interest

Transcript Name	Sequence ID No.
M78076_PEA_1_T2	74
M78076_PEA_1_T3	75
M78076_PEA_1_T5	76
M78076_PEA_1_T13	77
M78076_PEA_1_T15	78
M78076_PEA_1_T23	79
M78076_PEA_1_T26	80
M78076_PEA_1_T27	81
M78076_PEA_1_T28	82

Table 687 - Segments of interest

Segment Name	Sequence ID No.
M78076_PEA_1_node_0	659
M78076_PEA_1_node_10	660
M78076_PEA_1_node_15	661
M78076_PEA_1_node_18	662
M78076_PEA_1_node_20	663
M78076_PEA_1_node_24	664
M78076_PEA_1_node_26	665
M78076_PEA_1_node_29	666
M78076_PEA_1_node_32	667
M78076_PEA_1_node_35	668
M78076_PEA_1_node_37	669

M78076_PEA_1_node_46	670
M78076_PEA_1_node_47	671
M78076_PEA_1_node_54	672
M78076_PEA_1_node_1	673
M78076_PEA_1_node_2	674
M78076_PEA_1_node_3	675
M78076_PEA_1_node_6	676
M78076_PEA_1_node_7	677
M78076_PEA_1_node_12	678
M78076_PEA_1_node_22	679
M78076_PEA_1_node_27	680
M78076_PEA_1_node_30	681
M78076_PEA_1_node_31	682
M78076_PEA_1_node_34	683
M78076_PEA_1_node_36	684
M78076_PEA_1_node_41	685
M78076_PEA_1_node_42	686
M78076_PEA_1_node_43	687
M78076_PEA_1_node_45	688
M78076_PEA_1_node_49	689
M78076_PEA_1_node_50	690
M78076_PEA_1_node_51	691
M78076_PEA_1_node_52	692
M78076_PEA_1_node_53	693

Table 688 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
M78076_PEA_1_P3	1350	M78076_PEA_1_T2; M78076_PEA_1_T5

M78076_PEA_1_P4	1351	M78076_PEA_1_T3
M78076_PEA_1_P12	1352	M78076_PEA_1_T13
M78076_PEA_1_P14	1353	M78076_PEA_1_T15
M78076_PEA_1_P21	1354	M78076_PEA_1_T23
M78076_PEA_1_P24	1355	M78076_PEA_1_T26
M78076_PEA_1_P2	1356	M78076_PEA_1_T27
M78076_PEA_1_P25	1357	M78076_PEA_1_T28

These sequences are variants of the known protein Amyloid-like protein 1 precursor (SwissProt accession identifier APP1_HUMAN; known also according to the synonyms APLP; APLP-1), SEQ ID NO: 1439, referred to herein as the previously known protein.

Protein Amyloid-like protein 1 precursor is known or believed to have the following function(s): May play a role in postsynaptic function. The C-terminal gamma-secretase processed fragment, ALID1, activates transcription activation through APBB1 (Fe65) binding (By similarity). Couples to JIP signal transduction through C-terminal binding. May interact with cellular G-protein signaling pathways. Can regulate neurite outgrowth through binding to components of the extracellular matrix such as heparin and collagen I. The gamma-CTF peptide, C30, is a potent enhancer of neuronal apoptosis (By similarity). The sequence for protein Amyloid-like protein 1 precursor is given at the end of the application, as "Amyloid-like protein 1 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 689.

Table 689 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
48	A -> P

Protein Amyloid-like protein 1 precursor localization is believed to be Type I membrane protein. C-terminally processed in the Golgi complex.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: endocytosis; apoptosis; cell adhesion; neurogenesis; cell death, which

are annotation(s) related to Biological Process; protein binding; heparin binding, which are annotation(s) related to Molecular Function; and basement membrane; coated pit; integral membrane protein, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster M78076 features 9 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Amyloid-like protein 1 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein M78076_PEA_1_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T2. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P3 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P3, comprising a first amino acid sequence being at least 90 % homologous to
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKD corresponding to amino acids 1 - 517 of APP1_HUMAN, which also corresponds to amino acids 1 - 517 of

M78076_PEA_1_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GE corresponding to amino acids 518 - 519 of M78076_PEA_1_P3, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P3 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 690, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 690 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No

214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P3, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 691(given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 691- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	461
551	no	

Variant protein M78076_PEA_1_P3 is encoded by the following transcript(s): M78076_PEA_1_T2, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T2 is shown in bold; this coding portion starts at position 142 and ends at position 1698. The transcript also has the following SNPs as listed in Table 692 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 692 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
2146	G -> A	Yes
2224	C -> T	No
2362	C -> T	Yes
2513	A -> G	No
2656	C -> T	Yes

Variant protein M78076_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T3. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P4 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKG corresponding to amino acids 1 - 526 of APP1_HUMAN, which also corresponds to amino acids 1 - 526 of M78076_PEA_1_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ECLTVNPSLQIPLNP corresponding to amino acids 527 - 541 of M78076_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ECLTVNPSLQIPLNP in M78076_PEA_1_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 693, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 693 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No

270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P4, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 694(given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 694 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	461
551	no	

Variant protein M78076_PEA_1_P4 is encoded by the following transcript(s): M78076_PEA_1_T3, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T3 is shown in bold; this coding portion starts at position 142 and ends at position 1764. The transcript also has the following SNPs as listed in Table 695 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 695 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes

158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
1817	G -> A	Yes
2362	G -> A	Yes
2440	C -> T	No
2578	C -> T	Yes
2729	A -> G	No
2872	C -> T	Yes

Variant protein M78076_PEA_1_P12 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T13. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P12 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P12, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKG corresponding to amino acids 1 - 526 of APP1_HUMAN, which also corresponds to amino acids 1 - 526 of M78076_PEA_1_P12, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ECVCSKGFPFPLIGDSEG corresponding to amino acids 527 - 544 of M78076_PEA_1_P12, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P12, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ECVCSKGFPFPLIGDSEG in M78076_PEA_1_P12.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P12 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 696, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 696 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P12, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 697 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 697- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	461
551	no	

Variant protein M78076_PEA_1_P12 is encoded by the following transcript(s): M78076_PEA_1_T13, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T13 is shown in bold; this coding portion starts at position 142 and ends at position 1773. The transcript also has the following SNPs as listed in Table 698 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P12 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 698 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes

243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
1816	G -> A	Yes
1894	C -> T	No
2032	C -> T	Yes
2183	A -> G	No
2326	C -> T	Yes

Variant protein M78076_PEA_1_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T15. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the

relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P14 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P14, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD QNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGST EQDAASPEKEKMNPLEQYERKVNASVPRGFPFHSSEIQRDEL corresponding to amino acids 1 - 570 of APP1_HUMAN, which also corresponds to amino acids 1 - 570 of M78076_PEA_1_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRGGTAGYLGEETRGQRPGCDSQSHTGPSKKPSAPSPLPAGTSWDRGVP corresponding to amino acids 571 - 619 of M78076_PEA_1_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRGGTAGYLGEETRGQRPGCDSQSHTGPSKKPSAPSPLPAGTSWDRGVP in M78076_PEA_1_P14.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 699, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 699- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P14, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 700 (given according to

their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 700 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	461
551	yes	551

Variant protein M78076_PEA_1_P14 is encoded by the following transcript(s): M78076_PEA_1_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T15 is shown in bold; this coding portion starts at position 142 and ends at position 1998. The transcript also has the following SNPs as listed in Table 701(given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 701 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes

366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
2008	G -> A	Yes
2086	C -> T	No
2224	C -> T	Yes
2375	A -> G	No
2518	C -> T	Yes

Variant protein M78076_PEA_1_P21 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T23. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P21 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P21, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN E corresponding to amino acids 1 - 352 of APP1_HUMAN, which also corresponds to amino acids 1 - 352 of M78076_PEA_1_P21, and a second amino acid sequence being at least 90 % homologous to AERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMT LPKGSTEQDAASPEKEKMNPLEQYERKVNASVPRGFPFHSSEIQRDELAPAGTGVSREA VSGLLIMGAGGGSLIVLSMLLLRRKKPYGAISHGVVEVDPMLTLEEQQLRELQRHGYE NPTYRFLEERP corresponding to amino acids 406 - 650 of APP1_HUMAN, which also corresponds to amino acids 353 - 597 of M78076_PEA_1_P21, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of M78076_PEA_1_P21, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EA, having a structure as follows: a sequence starting from any of amino acid numbers 352-x to 352; and ending at any of amino acid numbers 353+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although both signal-peptide prediction programs agree that this protein has a signal peptide, both trans-membrane

region prediction programs predict that this protein has a trans-membrane region downstream of this signal peptide.

Variant protein M78076_PEA_1_P21 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 702, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P21 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 702 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes

The glycosylation sites of variant protein M78076_PEA_1_P21, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 703 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates

whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 703- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	408
551	yes	498

Variant protein M78076_PEA_1_P21 is encoded by the following transcript(s): M78076_PEA_1_T23, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T23 is shown in bold; this coding portion starts at position 142 and ends at position 1932. The transcript also has the following SNPs as listed in Table 704 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P21 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 704 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes

404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1239	G -> T	Yes
1264	C -> T	Yes
1728	G -> A	Yes
1806	C -> T	No
1944	C -> T	Yes
2095	A -> G	No
2238	C -> T	Yes

Variant protein M78076_PEA_1_P24 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T26. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P24 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P24, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL

CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN
EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL
ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQSLGLLD
QNPHLAQELRPQI corresponding to amino acids 1 - 481 of APP1_HUMAN, which also corresponds to amino acids 1 - 481 of M78076_PEA_1_P24, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RECLLPWLPLQISEGRS corresponding to amino acids 482 - 498 of M78076_PEA_1_P24, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P24, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RECLLPWLPLQISEGRS in M78076_PEA_1_P24.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P24 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 705, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P24 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 705 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P24, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 706 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 706 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	yes	461

551	no	

Variant protein M78076_PEA_1_P24 is encoded by the following transcript(s): M78076_PEA_1_T26, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T26 is shown in bold; this coding portion starts at position 142 and ends at position 1635. The transcript also has the following SNPs as listed in Table 707 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P24 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 707 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No

951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
2184	G -> A	Yes

Variant protein M78076_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T27. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P2 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQV corresponding to amino acids 1 - 449 of APP1_HUMAN, which also corresponds to amino acids 1 - 449 of M78076_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

LTSFQLPNAPLFLRRPRLRLFSCPLDPLSVSWTPSYPLNTASLPLPSLSAQLPDPETWTLT CCVFDPCFLALGFLLPPPSILCSVPWIFTAFPRIVFFFFFFLRQVLALSPRQESSVRSWLIAT STSWVQAILLPQPLE corresponding to amino acids 450 - 588 of M78076_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LTSFQLPNAPLFLRRPRLRLFSCPLDPLSVSWTPSYPLNTASLPLPSLSAQLPDPETWTLT CCVFDPCFLALGFLLPPPSILCSVPWIFTAFPRIVFFFFFFLRQVLALSPRQESSVRSWLIAT STSWVQAILLPQPLE in M78076_PEA_1_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although both signal-peptide prediction programs agree that this protein has a signal peptide, both trans-membrane region prediction programs predict that this protein has a trans-membrane region downstream of this signal peptide.

Variant protein M78076_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 708, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 708 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
4	A -> P	Yes

6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No
520	A -> S	Yes
546	F ->	Yes
564	S -> C	Yes

The glycosylation sites of variant protein M78076_PEA_1_P2, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 709 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 709 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	no	
551	no	

Variant protein M78076_PEA_1_P2 is encoded by the following transcript(s): M78076_PEA_1_T27, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript M78076_PEA_1_T27 is shown in bold; this coding portion starts at position 142 and ends at position 1905. The transcript also has the following SNPs as listed in Table 710 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 710 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes

1077	G -> A	Yes
1251	G ->	No
1398	G -> T	Yes
1423	C -> T	Yes
1500	C -> T	Yes
1699	G -> T	Yes
1725	G -> A	Yes
1777	T ->	Yes
1831	A -> T	Yes
2274	A -> G	Yes
2525	A -> G	Yes
2681	G -> A	Yes
3831	G -> A	Yes

Variant protein M78076_PEA_1_P25 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) M78076_PEA_1_T28. An alignment is given to the known protein (Amyloid-like protein 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between M78076_PEA_1_P25 and APP1_HUMAN:

1.An isolated chimeric polypeptide encoding for M78076_PEA_1_P25, comprising a first amino acid sequence being at least 90 % homologous to
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGL
CGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPME
RWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPEGCRFLHQERMDQCESSTRRHQ
EAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPG
SRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGV
DIYFGMPGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALN

EHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQADPPQAERVLL ALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQ corresponding to amino acids 1 - 448 of APP1_HUMAN, which also corresponds to amino acids 1 - 448 of M78076_PEA_1_P25, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PQNPNSQPRAAGSLEVIISHPFVRRLEILISPFQFQNSIPKNSQIVPAASPRGTSSP corresponding to amino acids 449 - 505 of M78076_PEA_1_P25, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of M78076_PEA_1_P25, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PQNPNSQPRAAGSLEVIISHPFVRRLEILISPFQFQNSIPKNSQIVPAASPRGTSSP in M78076_PEA_1_P25.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein M78076_PEA_1_P25 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 711, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P25 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 711 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?

4	A -> P	Yes
6	P -> H	Yes
13	R -> H	Yes
34	Q ->	No
38	G -> R	Yes
88	P -> R	Yes
124	R -> Q	Yes
127	S ->	No
145	F -> S	No
214	G -> R	No
214	G ->	No
262	Q ->	No
270	V ->	No
309	G -> E	Yes
370	Q ->	No

The glycosylation sites of variant protein M78076_PEA_1_P25, as compared to the known protein Amyloid-like protein 1 precursor, are described in Table 712 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 712- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
337	yes	337
461	no	
551	no	

Variant protein M78076_PEA_1_P25 is encoded by the following transcript(s): M78076_PEA_1_T28, for which the sequence(s) is/are given at the end of the application. The

coding portion of transcript M78076_PEA_1_T28 is shown in bold; this coding portion starts at position 142 and ends at position 1656. The transcript also has the following SNPs as listed in Table 713 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein M78076_PEA_1_P25 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 713 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
114	G ->	No
151	G -> C	Yes
158	C -> A	Yes
179	G -> A	Yes
219	A -> G	Yes
243	G ->	No
253	G -> A	Yes
315	A -> G	Yes
366	A -> G	Yes
404	C -> G	Yes
512	G -> A	Yes
522	C ->	No
522	C -> T	No
575	T -> C	No
781	G ->	No
781	G -> A	No
927	G ->	No
951	C ->	No
1067	G -> A	Yes
1077	G -> A	Yes
1251	G ->	No

1398	G -> T	Yes
1423	C -> T	Yes
1593	A -> G	No
1736	C -> T	Yes

As noted above, cluster M78076 features 35 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster M78076_PEA_1_node_0 according to the present invention is supported by 47 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 714 below describes the starting and ending position of this segment on each transcript.

Table 714 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1	160
M78076_PEA_1_T3	1	160
M78076_PEA_1_T5	1	160
M78076_PEA_1_T13	1	160
M78076_PEA_1_T15	1	160
M78076_PEA_1_T23	1	160
M78076_PEA_1_T26	1	160
M78076_PEA_1_T27	1	160
M78076_PEA_1_T28	1	160

Segment cluster M78076_PEA_1_node_10 according to the present invention is supported by 70 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 715 below describes the starting and ending position of this segment on each transcript.

Table 715 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	433	565
M78076_PEA_1_T3	433	565
M78076_PEA_1_T5	433	565
M78076_PEA_1_T13	433	565
M78076_PEA_1_T15	433	565
M78076_PEA_1_T23	433	565
M78076_PEA_1_T26	433	565
M78076_PEA_1_T27	433	565
M78076_PEA_1_T28	433	565

Segment cluster M78076_PEA_1_node_15 according to the present invention is supported by 74 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 716 below describes the starting and ending position of this segment on each transcript.

Table 716 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

M78076_PEA_1_T2	679	812
M78076_PEA_1_T3	679	812
M78076_PEA_1_T5	679	812
M78076_PEA_1_T13	679	812
M78076_PEA_1_T15	679	812
M78076_PEA_1_T23	679	812
M78076_PEA_1_T26	679	812
M78076_PEA_1_T27	679	812
M78076_PEA_1_T28	679	812

Segment cluster M78076_PEA_1_node_18 according to the present invention is supported by 95 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 717 below describes the starting and ending position of this segment on each transcript.

Table 717 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	813	991
M78076_PEA_1_T3	813	991
M78076_PEA_1_T5	813	991
M78076_PEA_1_T13	813	991
M78076_PEA_1_T15	813	991
M78076_PEA_1_T23	813	991
M78076_PEA_1_T26	813	991
M78076_PEA_1_T27	813	991
M78076_PEA_1_T28	813	991

Segment cluster M78076_PEA_1_node_20 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 718 below describes the starting and ending position of this segment on each transcript.

Table 718 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	992	1122
M78076_PEA_1_T3	992	1122
M78076_PEA_1_T5	992	1122
M78076_PEA_1_T13	992	1122
M78076_PEA_1_T15	992	1122
M78076_PEA_1_T23	992	1122
M78076_PEA_1_T26	992	1122
M78076_PEA_1_T27	992	1122
M78076_PEA_1_T28	992	1122

Segment cluster M78076_PEA_1_node_24 according to the present invention is supported by 105 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 719 below describes the starting and ending position of this segment on each transcript.

Table 719 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1198	1356
M78076_PEA_1_T3	1198	1356
M78076_PEA_1_T5	1198	1356
M78076_PEA_1_T13	1198	1356
M78076_PEA_1_T15	1198	1356
M78076_PEA_1_T26	1198	1356
M78076_PEA_1_T27	1198	1356
M78076_PEA_1_T28	1198	1356

Segment cluster M78076_PEA_1_node_26 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 720 below describes the starting and ending position of this segment on each transcript.

Table 720 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1357	1485
M78076_PEA_1_T3	1357	1485
M78076_PEA_1_T5	1357	1485
M78076_PEA_1_T13	1357	1485
M78076_PEA_1_T15	1357	1485
M78076_PEA_1_T23	1198	1326
M78076_PEA_1_T26	1357	1485
M78076_PEA_1_T27	1357	1485

M78076_PEA_1_T28	1357	1485

Segment cluster M78076_PEA_1_node_29 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T27. Table 721 below describes the starting and ending position of this segment on each transcript.

Table 721 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T27	1490	3132

Segment cluster M78076_PEA_1_node_32 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T26 and M78076_PEA_1_T27. Table 722 below describes the starting and ending position of this segment on each transcript.

Table 722 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T26	1586	2457
M78076_PEA_1_T27	3233	4104

Segment cluster M78076_PEA_1_node_35 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2 and M78076_PEA_1_T5. Table 723 below describes the starting and ending position of this segment on each transcript.

Table 723- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1694	1952
M78076_PEA_1_T5	1694	1952

Segment cluster M78076_PEA_1_node_37 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T3 and M78076_PEA_1_T5. Table 724 below describes the starting and ending position of this segment on each transcript.

Table 724 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T3	1718	2180
M78076_PEA_1_T5	1977	2439

Segment cluster M78076_PEA_1_node_46 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T15. Table 725 below describes the starting and ending position of this segment on each transcript.

Table 725 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T15	1852	1972

Segment cluster M78076_PEA_1_node_47 according to the present invention is supported by 155 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 726 below describes the starting and ending position of this segment on each transcript.

Table 726 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2111	2254
M78076_PEA_1_T3	2327	2470
M78076_PEA_1_T5	2586	2729
M78076_PEA_1_T13	1781	1924
M78076_PEA_1_T15	1973	2116
M78076_PEA_1_T23	1693	1836

Segment cluster M78076_PEA_1_node_54 according to the present invention is supported by 133 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23 and M78076_PEA_1_T28. Table 727 below describes the starting and ending position of this segment on each transcript.

Table 727 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2412	2715
M78076_PEA_1_T3	2628	2931
M78076_PEA_1_T5	2887	3190

M78076_PEA_1_T13	2082	2385
M78076_PEA_1_T15	2274	2577
M78076_PEA_1_T23	1994	2297
M78076_PEA_1_T28	1492	1795

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster M78076_PEA_1_node_1 according to the present invention is supported by 47 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 728 below describes the starting and ending position of this segment on each transcript.

Table 728- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	161	204
M78076_PEA_1_T3	161	204
M78076_PEA_1_T5	161	204
M78076_PEA_1_T13	161	204
M78076_PEA_1_T15	161	204
M78076_PEA_1_T23	161	204
M78076_PEA_1_T26	161	204
M78076_PEA_1_T27	161	204
M78076_PEA_1_T28	161	204

Segment cluster M78076_PEA_1_node_2 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5,

M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 729 below describes the starting and ending position of this segment on each transcript.

Table 729 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	205	224
M78076_PEA_1_T3	205	224
M78076_PEA_1_T5	205	224
M78076_PEA_1_T13	205	224
M78076_PEA_1_T15	205	224
M78076_PEA_1_T23	205	224
M78076_PEA_1_T26	205	224
M78076_PEA_1_T27	205	224
M78076_PEA_1_T28	205	224

Segment cluster M78076_PEA_1_node_3 according to the present invention is supported by 52 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 730 below describes the starting and ending position of this segment on each transcript.

Table 730 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	225	288
M78076_PEA_1_T3	225	288
M78076_PEA_1_T5	225	288

M78076_PEA_1_T13	225	288
M78076_PEA_1_T15	225	288
M78076_PEA_1_T23	225	288
M78076_PEA_1_T26	225	288
M78076_PEA_1_T27	225	288
M78076_PEA_1_T28	225	288

Segment cluster M78076_PEA_1_node_6 according to the present invention is supported by 59 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 731 below describes the starting and ending position of this segment on each transcript.

Table 731 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	289	370
M78076_PEA_1_T3	289	370
M78076_PEA_1_T5	289	370
M78076_PEA_1_T13	289	370
M78076_PEA_1_T15	289	370
M78076_PEA_1_T23	289	370
M78076_PEA_1_T26	289	370
M78076_PEA_1_T27	289	370
M78076_PEA_1_T28	289	370

Segment cluster M78076_PEA_1_node_7 according to the present invention is supported by 64 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 732 below describes the starting and ending position of this segment on each transcript.

Table 732 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	371	432
M78076_PEA_1_T3	371	432
M78076_PEA_1_T5	371	432
M78076_PEA_1_T13	371	432
M78076_PEA_1_T15	371	432
M78076_PEA_1_T23	371	432
M78076_PEA_1_T26	371	432
M78076_PEA_1_T27	371	432
M78076_PEA_1_T28	371	432

Segment cluster M78076_PEA_1_node_12 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 733 below describes the starting and ending position of this segment on each transcript.

Table 733- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	566	678
M78076_PEA_1_T3	566	678

M78076_PEA_1_T5	566	678
M78076_PEA_1_T13	566	678
M78076_PEA_1_T15	566	678
M78076_PEA_1_T23	566	678
M78076_PEA_1_T26	566	678
M78076_PEA_1_T27	566	678
M78076_PEA_1_T28	566	678

Segment cluster M78076_PEA_1_node_22 according to the present invention is supported by 92 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26, M78076_PEA_1_T27 and M78076_PEA_1_T28. Table 734 below describes the starting and ending position of this segment on each transcript.

Table 734 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1123	1197
M78076_PEA_1_T3	1123	1197
M78076_PEA_1_T5	1123	1197
M78076_PEA_1_T13	1123	1197
M78076_PEA_1_T15	1123	1197
M78076_PEA_1_T23	1123	1197
M78076_PEA_1_T26	1123	1197
M78076_PEA_1_T27	1123	1197
M78076_PEA_1_T28	1123	1197

Segment cluster M78076_PEA_1_node_27 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T27. Table 735 below describes the starting and ending position of this segment on each transcript.

Table 735 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T27	1486	1489

Segment cluster M78076_PEA_1_node_30 according to the present invention is supported by 90 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26 and M78076_PEA_1_T27. Table 736 below describes the starting and ending position of this segment on each transcript.

Table 736 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1486	1557
M78076_PEA_1_T3	1486	1557
M78076_PEA_1_T5	1486	1557
M78076_PEA_1_T13	1486	1557
M78076_PEA_1_T15	1486	1557
M78076_PEA_1_T23	1327	1398
M78076_PEA_1_T26	1486	1557
M78076_PEA_1_T27	3133	3204

Segment cluster M78076_PEA_1_node_31 according to the present invention is supported by 89 libraries. The number of libraries was determined as previously described. This

segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23, M78076_PEA_1_T26 and M78076_PEA_1_T27. Table 737 below describes the starting and ending position of this segment on each transcript.

Table 737- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1558	1585
M78076_PEA_1_T3	1558	1585
M78076_PEA_1_T5	1558	1585
M78076_PEA_1_T13	1558	1585
M78076_PEA_1_T15	1558	1585
M78076_PEA_1_T23	1399	1426
M78076_PEA_1_T26	1558	1585
M78076_PEA_1_T27	3205	3232

Segment cluster M78076_PEA_1_node_34 according to the present invention is supported by 103 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 738 below describes the starting and ending position of this segment on each transcript.

Table 738 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1586	1693
M78076_PEA_1_T3	1586	1693
M78076_PEA_1_T5	1586	1693

M78076_PEA_1_T13	1586	1693
M78076_PEA_1_T15	1586	1693
M78076_PEA_1_T23	1427	1534

Segment cluster M78076_PEA_1_node_36 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 739 below describes the starting and ending position of this segment on each transcript.

Table 739 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1953	1976
M78076_PEA_1_T3	1694	1717
M78076_PEA_1_T5	1953	1976
M78076_PEA_1_T13	1694	1717
M78076_PEA_1_T15	1694	1717
M78076_PEA_1_T23	1535	1558

Segment cluster M78076_PEA_1_node_41 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T3 and M78076_PEA_1_T5. Table 740 below describes the starting and ending position of this segment on each transcript.

Table 740 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T3	2181	2192
M78076_PEA_1_T5	2440	2451

Segment cluster M78076_PEA_1_node_42 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 741 below describes the starting and ending position of this segment on each transcript.

Table 741 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1977	1985
M78076_PEA_1_T3	2193	2201
M78076_PEA_1_T5	2452	2460
M78076_PEA_1_T15	1718	1726
M78076_PEA_1_T23	1559	1567

Segment cluster M78076_PEA_1_node_43 according to the present invention is supported by 110 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 742 below describes the starting and ending position of this segment on each transcript.

Table 742 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	1986	2047
M78076_PEA_1_T3	2202	2263
M78076_PEA_1_T5	2461	2522
M78076_PEA_1_T15	1727	1788
M78076_PEA_1_T23	1568	1629

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 743.

Table 743 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
M78076_0_7_0	lung malignant tumors	LUN

Segment cluster M78076_PEA_1_node_45 according to the present invention is supported by 132 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 744 below describes the starting and ending position of this segment on each transcript.

Table 744 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2048	2110
M78076_PEA_1_T3	2264	2326
M78076_PEA_1_T5	2523	2585
M78076_PEA_1_T13	1718	1780
M78076_PEA_1_T15	1789	1851
M78076_PEA_1_T23	1630	1692

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 745.

Table 745 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
M78076_0_7_0	lung malignant tumors	LUN

Segment cluster M78076_PEA_1_node_49 according to the present invention is supported by 129 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 746 below describes the starting and ending position of this segment on each transcript.

Table 746 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2255	2290
M78076_PEA_1_T3	2471	2506
M78076_PEA_1_T5	2730	2765
M78076_PEA_1_T13	1925	1960
M78076_PEA_1_T15	2117	2152
M78076_PEA_1_T23	1837	1872

Segment cluster M78076_PEA_1_node_50 according to the present invention is supported by 125 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 747 below describes the starting and ending position of this segment on each transcript.

Table 747 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2291	2329
M78076_PEA_1_T3	2507	2545
M78076_PEA_1_T5	2766	2804
M78076_PEA_1_T13	1961	1999
M78076_PEA_1_T15	2153	2191
M78076_PEA_1_T23	1873	1911

Segment cluster M78076_PEA_1_node_51 according to the present invention is supported by 123 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 748 below describes the starting and ending position of this segment on each transcript.

Table 748 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2330	2388
M78076_PEA_1_T3	2546	2604
M78076_PEA_1_T5	2805	2863
M78076_PEA_1_T13	2000	2058
M78076_PEA_1_T15	2192	2250
M78076_PEA_1_T23	1912	1970

Segment cluster M78076_PEA_1_node_52 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3,

M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15 and M78076_PEA_1_T23. Table 749 below describes the starting and ending position of this segment on each transcript.

Table 749 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2389	2405
M78076_PEA_1_T3	2605	2621
M78076_PEA_1_T5	2864	2880
M78076_PEA_1_T13	2059	2075
M78076_PEA_1_T15	2251	2267
M78076_PEA_1_T23	1971	1987

Segment cluster M78076_PEA_1_node_53 according to the present invention can be found in the following transcript(s): M78076_PEA_1_T2, M78076_PEA_1_T3, M78076_PEA_1_T5, M78076_PEA_1_T13, M78076_PEA_1_T15, M78076_PEA_1_T23 and M78076_PEA_1_T28. Table 750 below describes the starting and ending position of this segment on each transcript.

Table 750 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
M78076_PEA_1_T2	2406	2411
M78076_PEA_1_T3	2622	2627
M78076_PEA_1_T5	2881	2886
M78076_PEA_1_T13	2076	2081
M78076_PEA_1_T15	2268	2273
M78076_PEA_1_T23	1988	1993
M78076_PEA_1_T28	1486	1491

Variant protein alignment to the previously known protein:

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P3 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 5132.00

Escore: 0

Matching length: 517 Total length: 517

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                .         .         .         .         .
```

```
 51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
            .         .         .         .         .
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
            .         .         .         .         .
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
            .         .         .         .         .
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
            .         .         .         .         .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
            .         .         .         .         .
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
            .         .         .         .         .
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
            .         .         .         .         .
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
```

```
                .         .         .         .         .
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500

                .
501 GGSSEDKGGLQPPDSKD                                   517
    |||||||||||||||||
501 GGSSEDKGGLQPPDSKD                                   517
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P4 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 5223.00

Escore: 0

Matching length: 526 Total length: 526

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                 .         .         .         .         .
       1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                 .         .         .         .         .
      51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                 .         .         .         .         .
     101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                 .         .         .         .         .
     151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                 .         .         .         .         .
     201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
                 .         .         .         .         .
     251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                 .         .         .         .         .
     301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                 .         .         .         .         .
```

```
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
              .         .         .         .         .
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
              .         .         .         .         .
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
              .         .
501 GGSSEDKGGLQPPDSKDDTPMTLPKG                          526
    |||||||||||||||||||||||||
501 GGSSEDKGGLQPPDSKDDTPMTLPKG                          526
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P12 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 5223.00

Escore: 0

822

Matching length: 526 Total length: 526

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
          .         .         .         .         .
1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
  ||||||||||||||||||||||||||||||||||||||||||||||||||
1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
          .         .         .         .         .
51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
  ||||||||||||||||||||||||||||||||||||||||||||||||||
51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
          .         .         .         .         .
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
  ||||||||||||||||||||||||||||||||||||||||||||||||||
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
          .         .         .         .         .
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
  ||||||||||||||||||||||||||||||||||||||||||||||||||
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
          .         .         .         .         .
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
  ||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
          .         .         .         .         .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
```

```
                                                                                
                                                                                
          251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                                                                                
          301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                                                                                
          301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                                                                                
          351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                                                                                
          351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                                                                                
          401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
                                                                                
          401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
                                                                                
          451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
                                                                                
          451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
                                                                                
          501 GGSSEDKGGLQPPDSKDDTPMTLPKG                          526
                                                                                
          501 GGSSEDKGGLQPPDSKDDTPMTLPKG                          526
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P14 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 5672.00

Escore: 0

Matching length: 575 Total length: 575

Matching Percent Similarity: 99.48 Matching Percent Identity: 99.48

Total Percent Similarity: 99.48 Total Percent Identity: 99.48

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                .         .         .         .         .
       51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                .         .         .         .         .
      101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                .         .         .         .         .
      151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                .         .         .         .         .
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
                .         .         .         .         .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                .         .         .         .         .
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                .         .         .         .         .
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                .         .         .         .         .
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
                .         .         .         .         .
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
                .         .         .         .         .
501 GGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKV 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 GGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKV 550
                .         .
551 NASVPRGFPFHSSEIQRDELVRGGT                           575
```

826

||||||||||||||||||| ||

551 NASVPRGFPFHSSEIQRDELAPAGT 575

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P21 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 5822.00

Escore: 0

Matching length: 597 Total length: 650

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 91.85 Total Percent Identity: 91.85

Gaps: 1

Alignment:

. . . . .

1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50

|||||||||||||||||||||||||||||||||||||||||||||||||||

1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50

```
                .         .         .         .         .
 51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
 51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                .         .         .         .         .
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                .         .         .         .         .
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                .         .         .         .         .
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
                .         .         .         .         .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                .         .         .         .         .
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                .         .         .         .         .
351 NE..................................................  352
    ||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                .         .         .         .         .
353 .....AERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 397
          |||||||||||||||||||||||||||||||||||||||||||||
```

```
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
                .       .       .       .       .
398 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 447
    |||||||||||||||||||||||||||||||||||||||||||||||||||
451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAP 500
                .       .       .       .       .
448 GGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKV 497
    |||||||||||||||||||||||||||||||||||||||||||||||||||
501 GGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKV 550
                .       .       .       .       .
498 NASVPRGFPFHSSEIQRDELAPAGTGVSREAVSGLLIMGAGGGSLIVLSM 547
    |||||||||||||||||||||||||||||||||||||||||||||||||||
551 NASVPRGFPFHSSEIQRDELAPAGTGVSREAVSGLLIMGAGGGSLIVLSM 600
                .       .       .       .       .
548 LLLRRKKPYGAISHGVVEVDPMLTLEEQQLRELQRHGYENPTYRFLEERP 597
    |||||||||||||||||||||||||||||||||||||||||||||||||||
601 LLLRRKKPYGAISHGVVEVDPMLTLEEQQLRELQRHGYENPTYRFLEERP 650
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P24 x APP1_HUMAN ..

Alignment segment 1/1:

829

Quality: 4791.00

Escore: 0

Matching length: 485 Total length: 485

Matching Percent Similarity: 99.79 Matching Percent Identity: 99.59

Total Percent Similarity: 99.79 Total Percent Identity: 99.59

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                .         .         .         .         .
     51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                .         .         .         .         .
    101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                .         .         .         .         .
    151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                .         .         .         .         .
    201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
```

```
                .         .         .         .         .
      251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                .         .         .         .         .
      301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                .         .         .         .         .
      351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                .         .         .         .         .
      401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
          |||||||||||||||||||||||||||||||||||||||||||||||||||
      401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450
                .         .         .
      451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIRECL               485
          |||||||||||||||||||||||||||||:| |
      451 THLQVIEERVNQSLGLLDQNPHLAQELRPQIQELL               485
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P2 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 4474.00
Escore: 0
Matching length: 454 Total length: 454
Matching Percent Similarity: 99.56 Matching Percent Identity: 99.34
Total Percent Similarity: 99.56 Total Percent Identity: 99.34
Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                .         .         .         .         .
     51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                .         .         .         .         .
    101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                .         .         .         .         .
    151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                .         .         .         .         .
```

```
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
          .         .         .         .         .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
          .         .         .         .         .
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
          .         .         .         .         .
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
          .         .         .         .         .
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVL 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVH 450

451 TSFQ                                               454
    | :|
451 THLQ                                               454
```

Sequence name: APP1_HUMAN

Sequence documentation:

Alignment of: M78076_PEA_1_P25 x APP1_HUMAN ..

Alignment segment 1/1:

Quality: 4455.00

Escore: 0

Matching length: 448 Total length: 448

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEA 50
                .         .         .         .         .
       51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 PGSAQVAGLCGRLTLHRDLRTGRWEPDPQRSRRCLRDPQRVLEYCRQMYP 100
                .         .         .         .         .
      101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      101 ELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEAL 150
                .         .         .         .         .
      151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
```

```
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
151 LVPEGCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSD 200
                                                                                .          .          .          .          .
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
201 RFRGVEYVCCPPPGTPDPSGTAVGDPSTRSWPPGSRVEGAEDEEEEESFP 250
                                                                                .          .          .          .          .
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
251 QPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGM 300
                                                                                .          .          .          .          .
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 PGEISEHEGFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQAL 350
                                                                                .          .          .          .          .
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NEHFQSILQTLEEQVSGERQRLVETHATRVIALINDQRRAALEGFLAALQ 400
                                                                                .          .          .          .
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQ    448
                                                                    |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
401 ADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQ    448
```

DESCRIPTION FOR CLUSTER T99080

Cluster T99080 features 14 transcript(s) and 11 segment(s) of interest, the names for which are given in Tables 751 and 752, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 753.

Table 751 - Transcripts of interest

Transcript Name	Sequence ID No.
T99080_PEA_4_T0	83
T99080_PEA_4_T2	84
T99080_PEA_4_T4	85
T99080_PEA_4_T6	86
T99080_PEA_4_T9	87
T99080_PEA_4_T10	88
T99080_PEA_4_T11	89
T99080_PEA_4_T13	90
T99080_PEA_4_T14	91
T99080_PEA_4_T17	92
T99080_PEA_4_T18	93
T99080_PEA_4_T19	94
T99080_PEA_4_T20	95
T99080_PEA_4_T21	96

Table 752 - Segments of interest

Segment Name	Sequence ID No.
T99080_PEA_4_node_1	695
T99080_PEA_4_node_6	696
T99080_PEA_4_node_11	697
T99080_PEA_4_node_19	698
T99080_PEA_4_node_20	699
T99080_PEA_4_node_3	700
T99080_PEA_4_node_5	701
T99080_PEA_4_node_8	702
T99080_PEA_4_node_13	703
T99080_PEA_4_node_15	704
T99080_PEA_4_node_18	705

Table 753 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
T99080_PEA_4_P1	1358	T99080_PEA_4_T0
T99080_PEA_4_P2	1359	T99080_PEA_4_T2
T99080_PEA_4_P5	1360	T99080_PEA_4_T6
T99080_PEA_4_P8	1361	T99080_PEA_4_T9
T99080_PEA_4_P9	1362	T99080_PEA_4_T10
T99080_PEA_4_P10	1363	T99080_PEA_4_T11
T99080_PEA_4_P12	1364	T99080_PEA_4_T14
T99080_PEA_4_P13	1365	T99080_PEA_4_T17
T99080_PEA_4_P14	1366	T99080_PEA_4_T18
T99080_PEA_4_P15	1367	T99080_PEA_4_T19
T99080_PEA_4_P16	1368	T99080_PEA_4_T20
T99080_PEA_4_P17	1369	T99080_PEA_4_T21

These sequences are variants of the known protein Acylphosphatase, organ-common type isozyme (SwissProt accession identifier ACYO_HUMAN; known also according to the synonyms EC 3.6.1.7; Acylphosphate phosphohydrolase; Acylphosphatase, erythrocyte isozyme), SEQ ID NO: 1440, referred to herein as the previously known protein.

The sequence for protein Acylphosphatase, organ-common type isozyme is given at the end of the application, as "Acylphosphatase, organ-common type isozyme amino acid sequence". Known polymorphisms for this sequence are as shown in Table 754.

Table 754 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
19	G -> R

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: phosphate metabolism, which are annotation(s) related to Biological Process; and acylphosphatase, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster T99080 features 14 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Acylphosphatase, organ-common type isozyme. A description of each variant protein according to the present invention is now provided.

Variant protein T99080_PEA_4_P1 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T0. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P1 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 755, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 755 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P1 is encoded by the following transcript(s): T99080_PEA_4_T0, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T0 is shown in bold; this coding portion starts at position 226 and ends at position 411. The transcript also has the following SNPs as listed in Table 756 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P1 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 756 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes
1293	G -> C	Yes
2034	A -> G	Yes
2114	A -> C	Yes
2153	-> A	No

Variant protein T99080_PEA_4_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T2. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P2 is encoded by the following transcript(s): T99080_PEA_4_T2, for which the sequence(s) is/are given at the end of the application. The

coding portion of transcript T99080_PEA_4_T2 is shown in bold; this coding portion starts at position 1 and ends at position 192. The transcript also has the following SNPs as listed in Table 757 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 757- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
1074	G -> C	Yes
1815	A -> G	Yes
1895	A -> C	Yes
1934	-> A	No

Variant protein T99080_PEA_4_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T6. An alignment is given to the known protein (Acylphosphatase, organ-common type isozyme) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T99080_PEA_4_P5 and ACYO_HUMAN_V1 (SEQ ID NO: 1441):

1.An isolated chimeric polypeptide encoding for T99080_PEA_4_P5, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MPASARLAGAGLLLAFLRALGCAGRAPGLS corresponding to amino acids 1 - 30 of T99080_PEA_4_P5, and a second amino acid sequence being at least 90 % homologous to

MAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQGQLQGPIS KVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK corresponding to amino acids 1 - 99 of ACYO_HUMAN_V1, which also corresponds to amino acids 31 - 129 of T99080_PEA_4_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of T99080_PEA_4_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MPASARLAGAGLLLAFLRALGCAGRAPGLS of T99080_PEA_4_P5.

It should be noted that the known protein sequence (ACYO_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ACYO_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 758 - Changes to ACYO_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 759, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P5

sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 759 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P5 is encoded by the following transcript(s): T99080_PEA_4_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T6 is shown in bold; this coding portion starts at position 226 and ends at position 612. The transcript also has the following SNPs as listed in Table 760 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 760 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes
697	A -> G	Yes
777	A -> C	Yes
816	-> A	No

Variant protein T99080_PEA_4_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T9. An alignment is given to the known protein (Acylphosphatase, organ-common type isozyme) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief

description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T99080_PEA_4_P8 and ACYO_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for T99080_PEA_4_P8, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence M corresponding to amino acids 1 - 1 of T99080_PEA_4_P8, and a second amino acid sequence being at least 90 % homologous to QAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHIDKANFNNE KVILKLDYSDFQIVK corresponding to amino acids 28 - 99 of ACYO_HUMAN_V1, which also corresponds to amino acids 2 - 73 of T99080_PEA_4_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

It should be noted that the known protein sequence (ACYO_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ACYO_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 761 - Changes to ACYO_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein T99080_PEA_4_P8 is encoded by the following transcript(s): T99080_PEA_4_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T9 is shown in bold; this coding portion starts at position 162 and ends at position 380. The transcript also has the following SNPs as listed in Table 762 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 762 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
465	A -> G	Yes
545	A -> C	Yes
584	-> A	No

Variant protein T99080_PEA_4_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T10. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P9 is encoded by the following transcript(s): T99080_PEA_4_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T10 is shown in bold; this coding portion starts at position 1 and ends at position 261. The transcript also has the following SNPs as listed in Table

763 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 763 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
557	A -> G	Yes
637	A -> C	Yes
676	-> A	No

Variant protein T99080_PEA_4_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T11. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P10 is encoded by the following transcript(s): T99080_PEA_4_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T11 is shown in bold; this coding portion starts at position 1 and ends at position 240. The transcript also has the following SNPs as listed in Table 764 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 764 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
269	G -> T	Yes
592	A -> G	Yes
672	A -> C	Yes
711	-> A	No

Variant protein T99080_PEA_4_P12 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T14. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P12 is encoded by the following transcript(s): T99080_PEA_4_T14, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T14 is shown in bold; this coding portion starts at position 1 and ends at position 282.

Variant protein T99080_PEA_4_P13 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T17. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although it is a partial protein, because both trans-membrane region prediction programs predict that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P13 is encoded by the following transcript(s): T99080_PEA_4_T17, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T17 is shown in bold; this coding portion starts at position 1 and ends at position 207.

Variant protein T99080_PEA_4_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T18. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 765, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 765 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P14 is encoded by the following transcript(s): T99080_PEA_4_T18, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T18 is shown in bold; this coding portion starts at

position 226 and ends at position 480. The transcript also has the following SNPs as listed in Table 766 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 766 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes
776	A -> G	Yes
856	A -> C	Yes
895	-> A	No

Variant protein T99080_PEA_4_P15 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T19. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P15 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 767, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 767 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P15 is encoded by the following transcript(s): T99080_PEA_4_T19, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T19 is shown in bold; this coding portion starts at position 226 and ends at position 459. The transcript also has the following SNPs as listed in Table 768 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 768 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes
488	G -> T	Yes
811	A -> G	Yes
891	A -> C	Yes
930	-> A	No

Variant protein T99080_PEA_4_P16 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T20. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither

trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P16 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 769, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 769 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P16 is encoded by the following transcript(s): T99080_PEA_4_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T20 is shown in bold; this coding portion starts at position 226 and ends at position 501. The transcript also has the following SNPs as listed in Table 770 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 770 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes

Variant protein T99080_PEA_4_P17 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T99080_PEA_4_T21. The location of the variant protein was determined according to results

from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T99080_PEA_4_P17 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 771, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 771 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
23	A -> V	Yes

Variant protein T99080_PEA_4_P17 is encoded by the following transcript(s): T99080_PEA_4_T21, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T99080_PEA_4_T21 is shown in bold; this coding portion starts at position 226 and ends at position 426. The transcript also has the following SNPs as listed in Table 772 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T99080_PEA_4_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 772 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
293	C -> T	Yes

As noted above, cluster T99080 features 11 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster T99080_PEA_4_node_1 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T0, T99080_PEA_4_T6, T99080_PEA_4_T13, T99080_PEA_4_T18, T99080_PEA_4_T19, T99080_PEA_4_T20 and T99080_PEA_4_T21. Table 773 below describes the starting and ending position of this segment on each transcript.

Table 773 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T0	1	307
T99080_PEA_4_T6	1	307
T99080_PEA_4_T13	1	307
T99080_PEA_4_T18	1	307
T99080_PEA_4_T19	1	307
T99080_PEA_4_T20	1	307
T99080_PEA_4_T21	1	307

Segment cluster T99080_PEA_4_node_6 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T17 and T99080_PEA_4_T21. Table 774 below describes the starting and ending position of this segment on each transcript.

Table 774 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T17	181	627
T99080_PEA_4_T21	400	846

Segment cluster T99080_PEA_4_node_11 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T14 and T99080_PEA_4_T20. Table 775 below describes the starting and ending position of this segment on each transcript.

Table 775 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T14	260	782
T99080_PEA_4_T20	479	1001

Segment cluster T99080_PEA_4_node_19 according to the present invention is supported by 59 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T0, T99080_PEA_4_T2 and T99080_PEA_4_T4. Table 776 below describes the starting and ending position of this segment on each transcript.

Table 776 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T0	449	1736
T99080_PEA_4_T2	230	1517
T99080_PEA_4_T4	78	1365

Segment cluster T99080_PEA_4_node_20 according to the present invention is supported by 98 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T0, T99080_PEA_4_T2, T99080_PEA_4_T4, T99080_PEA_4_T6, T99080_PEA_4_T9, T99080_PEA_4_T10, T99080_PEA_4_T11, T99080_PEA_4_T13, T99080_PEA_4_T18 and T99080_PEA_4_T19. Table 777 below describes the starting and ending position of this segment on each transcript.

Table 777 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T0	1737	2175
T99080_PEA_4_T2	1518	1956
T99080_PEA_4_T4	1366	1804
T99080_PEA_4_T6	400	838
T99080_PEA_4_T9	168	606
T99080_PEA_4_T10	260	698
T99080_PEA_4_T11	295	733
T99080_PEA_4_T13	308	746
T99080_PEA_4_T18	479	917
T99080_PEA_4_T19	514	952

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster T99080_PEA_4_node_3 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T2, T99080_PEA_4_T9, T99080_PEA_4_T10, T99080_PEA_4_T11, T99080_PEA_4_T14 and T99080_PEA_4_T17. Table 778 below describes the starting and ending position of this segment on each transcript.

Table 778 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T2	1	88
T99080_PEA_4_T9	1	88
T99080_PEA_4_T10	1	88
T99080_PEA_4_T11	1	88
T99080_PEA_4_T14	1	88
T99080_PEA_4_T17	1	88

Segment cluster T99080_PEA_4_node_5 according to the present invention is supported by 57 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T0, T99080_PEA_4_T2, T99080_PEA_4_T6, T99080_PEA_4_T10, T99080_PEA_4_T11, T99080_PEA_4_T14, T99080_PEA_4_T17, T99080_PEA_4_T18, T99080_PEA_4_T19, T99080_PEA_4_T20 and T99080_PEA_4_T21. Table 779 below describes the starting and ending position of this segment on each transcript.

Table 779 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T0	308	399
T99080_PEA_4_T2	89	180
T99080_PEA_4_T6	308	399
T99080_PEA_4_T10	89	180
T99080_PEA_4_T11	89	180
T99080_PEA_4_T14	89	180
T99080_PEA_4_T17	89	180
T99080_PEA_4_T18	308	399

T99080_PEA_4_T19	308	399
T99080_PEA_4_T20	308	399
T99080_PEA_4_T21	308	399

Segment cluster T99080_PEA_4_node_8 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T9, T99080_PEA_4_T10, T99080_PEA_4_T14, T99080_PEA_4_T18 and T99080_PEA_4_T20. Table 780 below describes the starting and ending position of this segment on each transcript.

Table 780 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T9	89	167
T99080_PEA_4_T10	181	259
T99080_PEA_4_T14	181	259
T99080_PEA_4_T18	400	478
T99080_PEA_4_T20	400	478

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 781.

Table 781 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
T99080_0_0_58896	lung malignant tumors	LUN

Segment cluster T99080_PEA_4_node_13 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T4. Table 782 below describes the starting and ending position of this segment on each transcript.

Table 782 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T4	1	77

Segment cluster T99080_PEA_4_node_15 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T11 and T99080_PEA_4_T19. Table 783 below describes the starting and ending position of this segment on each transcript.

Table 783 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T11	181	294
T99080_PEA_4_T19	400	513

Segment cluster T99080_PEA_4_node_18 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T99080_PEA_4_T0 and T99080_PEA_4_T2. Table 784 below describes the starting and ending position of this segment on each transcript.

Table 784 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T99080_PEA_4_T0	400	448

T99080_PEA_4_T2	181	229

Variant protein alignment to the previously known protein:

Sequence name: ACYO_HUMAN_V1

Sequence documentation:

Alignment of: T99080_PEA_4_P5 x ACYO_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 973.00

Escore: 0

Matching length: 99 Total length: 99

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      31 MAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQG 80
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

858

```
 1 MAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQG 50
                .         .         .         .
81 QLQGPISKVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK 129
   |||||||||||||||||||||||||||||||||||||||||||||||||||
51 QLQGPISKVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK 99
```

Sequence name: ACYO_HUMAN_V1

Sequence documentation:

Alignment of: T99080_PEA_4_P8 x ACYO_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 711.00

Escore: 0

Matching length: 72 Total length: 72

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

```
                                                    859
         2 QAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHID 51
           ||||||||||||||||||||||||||||||||||||||||||||||||||
        28 QAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHID 77
                           .         .
        52 KANFNNEKVILKLDYSDFQIVK                              73
           ||||||||||||||||||||||
        78 KANFNNEKVILKLDYSDFQIVK                              99
```

DESCRIPTION FOR CLUSTER T08446

Cluster T08446 features 2 transcript(s) and 36 segment(s) of interest, the names for which are given in Tables 785 and 786, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 787.

Table 785 - Transcripts of interest

Transcript Name	Sequence ID No.
T08446_PEA_1_T2	97
T08446_PEA_1_T22	98

Table 786 - Segments of interest

Segment Name	Sequence ID No.
T08446_PEA_1_node_2	706
T08446_PEA_1_node_9	707
T08446_PEA_1_node_15	708
T08446_PEA_1_node_17	709
T08446_PEA_1_node_25	710
T08446_PEA_1_node_29	711
T08446_PEA_1_node_38	712
T08446_PEA_1_node_43	713
T08446_PEA_1_node_51	714

T08446_PEA_1_node_52	715
T08446_PEA_1_node_55	716
T08446_PEA_1_node_57	717
T08446_PEA_1_node_59	718
T08446_PEA_1_node_62	719
T08446_PEA_1_node_63	720
T08446_PEA_1_node_3	721
T08446_PEA_1_node_5	722
T08446_PEA_1_node_7	723
T08446_PEA_1_node_12	724
T08446_PEA_1_node_13	725
T08446_PEA_1_node_19	726
T08446_PEA_1_node_21	727
T08446_PEA_1_node_23	728
T08446_PEA_1_node_27	729
T08446_PEA_1_node_32	730
T08446_PEA_1_node_34	731
T08446_PEA_1_node_45	732
T08446_PEA_1_node_46	733
T08446_PEA_1_node_48	734
T08446_PEA_1_node_54	735
T08446_PEA_1_node_58	736
T08446_PEA_1_node_60	737
T08446_PEA_1_node_61	738
T08446_PEA_1_node_64	739
T08446_PEA_1_node_65	740
T08446_PEA_1_node_66	741

Table 787 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
T08446_PEA_1_P18	1370	T08446_PEA_1_T2
T08446_PEA_1_P19	1371	T08446_PEA_1_T22

These sequences are variants of the known protein Sorting nexin 26 (SwissProt accession identifier SNXQ_HUMAN), SEQ ID NO: 1442, referred to herein as the previously known protein.

Protein Sorting nexin 26 is known or believed to have the following function(s): May be involved in several stages of intracellular trafficking (By similarity). The sequence for protein Sorting nexin 26 is given at the end of the application, as "Sorting nexin 26 amino acid sequence".

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: intracellular protein traffic, which are annotation(s) related to Biological Process; and protein transporter, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

As noted above, cluster T08446 features 2 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Sorting nexin 26. A description of each variant protein according to the present invention is now provided.

Variant protein T08446_PEA_1_P18 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T08446_PEA_1_T2. An alignment is given to the known protein (Sorting nexin 26) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T08446_PEA_1_P18 and SNXQ_HUMAN:

1.An isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 90 % homologous to

MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWME corresponding to amino acids 1 - 185 of SNXQ_HUMAN, which also
corresponds to amino acids 1 - 185 of T08446_PEA_1_P18, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSW
WRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA
GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVV
DGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLY
GKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNL
AIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPA
GRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEK
QRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLS
SQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGL
GALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAP
PAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTP
ALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLR
PGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEP
LGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQ
LRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQP
SSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGM
LGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPA
SSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY
GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC corresponding to amino acids 186 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the

sequence
LDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSW
WRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA
GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVV
DGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLY
GKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNL
AIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPA
GRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEK
QRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLS
SQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGL
GALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAP
PAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTP
ALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLR
PGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEP
LGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQ
LRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQP
SSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGM
LGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPA
SSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY
GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC in
T08446_PEA_1_P18.

Comparison report between T08446_PEA_1_P18 and Q9NT23 (SEQ ID NO: 1443):

1.An isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence
MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM

PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRV corresponding to amino acids 1 - 443 of
T08446_PEA_1_P18, a second amino acid sequence being at least 90 % homologous to
HDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVG
MGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGSCPSTR
LLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALGRG
PSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHS
SSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGLGALS GSPSHRTSAWLDDG
DELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRG
PTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHLIPLLLRGA
EAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALA
LAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAW
VPGPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSP
CSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDR
GENLYYEIGASEGSPYSG corresponding to amino acids 1 - 674 of Q9NT23, which also
corresponds to amino acids 444 - 1117 of T08446_PEA_1_P18, a bridging amino acid P
corresponding to amino acid 1118 of T08446_PEA_1_P18, and a third amino acid sequence
being at least 90 % homologous to
TRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPAR
RPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHR
VPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHS
EGQTRSYC corresponding to amino acids 676 - 862 of Q9NT23, which also corresponds to
amino acids 1119 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence,
second amino acid sequence, bridging amino acid and third amino acid sequence are contiguous
and in a sequential order.

2.An isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a
polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,
more preferably at least about 90% and most preferably at least about 95% homologous to the
sequence

865

MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM
PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRV of T08446_PEA_1_P18.

Comparison report between T08446_PEA_1_P18 and Q96CP3 (SEQ ID NO: 1444):

1.An isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM
PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANT
SMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTF
TSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAER
RKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRS
AKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSS
SESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDP
APPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGA
PASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLP
PPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS
LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRG

corresponding to amino acids 1 - 1010 of T08446_PEA_1_P18, and a second amino acid sequence being at least 90 % homologous to

LRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPG
LYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPP
DRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNL
ALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPL
LLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC

corresponding to amino acids 1 - 295 of Q96CP3, which also corresponds to amino acids 1011 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG
PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY
DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQMLVPLLLQYLETLSGLVDSNLNC
GPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIVSVIDM
PPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAV
PRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSE
FIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVSSLCKLYFRELPNP
LLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLLRHLARMARHSANT
SMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTF
TSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAER
RKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRS
AKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSS
SESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTPGDP
APPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGA
PASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLP
PPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS

LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPMGTSRRG of T08446_PEA_1_P18.

Comparison report between T08446_PEA_1_P18 and BAC86902 (SEQ ID NO: 1445):

1.An isolated chimeric polypeptide encoding for T08446_PEA_1_P18, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY DDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ corresponding to amino acids 1 - 154 of T08446_PEA_1_P18, a second amino acid sequence being at least 90 % homologous to MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVI KRYTAQAPDELSFEVGDIVSVIDMPPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPG LKADADGPPCGIPAPQGISSLTSAVPRPRGKLAGLLRTFMRSRPSRQRLRQRGILRQRVF GCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPEL SGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLP PPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGAAAFR EVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQ ARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALGRGPSVPRKKP LPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVG PAPAGSCESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPR CLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAA LDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQ QEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVA EQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPR QQSDGSLLRSQRPMGTSRRGLRGPA corresponding to amino acids 1 - 861 of BAC86902, which also corresponds to amino acids 155 - 1015 of T08446_PEA_1_P18, a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence QVSAQLRAGGGGRDAPEAAAQSPCSVPS corresponding to amino acids 1016 - 1043 of

T08446_PEA_1_P18, a fourth amino acid sequence being at least 90 % homologous to QVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLY YEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPP DHLGYS corresponding to amino acids 862 - 989 of BAC86902, which also corresponds to amino acids 1044 - 1171 of T08446_PEA_1_P18, and a fifth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence APQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAP WGPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYP TPSWSLHSEGQTRSYC corresponding to amino acids 1172 - 1305 of T08446_PEA_1_P18, wherein said first amino acid sequence, second amino acid sequence, third amino acid sequence, fourth amino acid sequence and fifth amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRG PFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSY DDFRSLDAHLHRCIFDRRFSCLP ELPPPPEGARAAQ of T08446_PEA_1_P18.

3.An isolated polypeptide encoding for an edge portion of T08446_PEA_1_P18, comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence encoding for QVSAQLRAGGGGRDAPEAAAQSPCSVPS, corresponding to T08446_PEA_1_P18.

4.An isolated polypeptide encoding for a tail of T08446_PEA_1_P18, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence APQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAP

WGPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYP TPSWSLHSEGQTRSYC in T08446_PEA_1_P18.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T08446_PEA_1_P18 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 788, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T08446_PEA_1_P18 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 788 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
714	S -> C	Yes
1000	S -> N	No
1273	R -> S	No
1274	N -> H	No

Variant protein T08446_PEA_1_P18 is encoded by the following transcript(s): T08446_PEA_1_T2, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T08446_PEA_1_T2 is shown in bold; this coding portion starts at position 228 and ends at position 4142. The transcript also has the following SNPs as listed in Table 789 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein T08446_PEA_1_P18 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 789 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
212	G -> A	Yes
431	C -> T	Yes
809	C -> T	Yes
1547	G -> A	Yes
2368	C -> G	Yes
3226	G -> A	No
3284	C -> G	Yes
3377	C -> T	Yes
4046	A -> C	No
4047	A -> C	No

Variant protein T08446_PEA_1_P19 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T08446_PEA_1_T22. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein T08446_PEA_1_P19 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 790, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T08446_PEA_1_P19

sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 790 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
194	D -> G	Yes

Variant protein T08446_PEA_1_P19 is encoded by the following transcript(s): T08446_PEA_1_T22, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T08446_PEA_1_T22 is shown in bold; this coding portion starts at position 228 and ends at position 965. The transcript also has the following SNPs as listed in Table 791 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T08446_PEA_1_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 791 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
212	G -> A	Yes
431	C -> T	Yes
808	A -> G	Yes

As noted above, cluster T08446 features 36 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster T08446_PEA_1_node_2 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 792 below describes the starting and ending position of this segment on each transcript.

Table 792 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1	287
T08446_PEA_1_T22	1	287

Segment cluster T08446_PEA_1_node_9 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 793 below describes the starting and ending position of this segment on each transcript.

Table 793 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	552	689
T08446_PEA_1_T22	552	689

Segment cluster T08446_PEA_1_node_15 according to the present invention is supported by 0 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T22. Table 794 below describes the starting and ending position of this segment on each transcript.

Table 794 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T22	829	968

Segment cluster T08446_PEA_1_node_17 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 794 below describes the starting and ending position of this segment on each transcript.

Table 794 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	783	905

Segment cluster T08446_PEA_1_node_25 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 12 below describes the starting and ending position of this segment on each transcript.

Table 12 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1111	1263

Segment cluster T08446_PEA_1_node_29 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 795 below describes the starting and ending position of this segment on each transcript.

Table 795 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1367	1511

Segment cluster T08446_PEA_1_node_38 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 796 below describes the starting and ending position of this segment on each transcript.

Table 796 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1703	1848

Segment cluster T08446_PEA_1_node_43 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 797 below describes the starting and ending position of this segment on each transcript.

Table 797 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1849	2002

Segment cluster T08446_PEA_1_node_51 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 798 below describes the starting and ending position of this segment on each transcript.

Table 798 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

T08446_PEA_1_T2	2224	2571

Segment cluster T08446_PEA_1_node_52 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 799 below describes the starting and ending position of this segment on each transcript.

Table 799 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2572	2694

Segment cluster T08446_PEA_1_node_55 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 800 below describes the starting and ending position of this segment on each transcript.

Table 800 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2707	2883

Segment cluster T08446_PEA_1_node_57 according to the present invention is supported by 37 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 801 below describes the starting and ending position of this segment on each transcript.

Table 801 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2884	3275

Segment cluster T08446_PEA_1_node_59 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 802 below describes the starting and ending position of this segment on each transcript.

Table 802 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3360	3670

Segment cluster T08446_PEA_1_node_62 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 803 below describes the starting and ending position of this segment on each transcript.

Table 803 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3783	3988

Segment cluster T08446_PEA_1_node_63 according to the present invention is supported by 64 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 804 below describes the starting and ending position of this segment on each transcript.

Table 804 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3989	4414

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster T08446_PEA_1_node_3 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 805 below describes the starting and ending position of this segment on each transcript.

Table 805 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	288	385
T08446_PEA_1_T22	288	385

Segment cluster T08446_PEA_1_node_5 according to the present invention is supported by 17 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 806 below describes the starting and ending position of this segment on each transcript.

Table 806 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	386	470
T08446_PEA_1_T22	386	470

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 807.

Table 807 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
T08446_0_9_0	lung malignant tumors	LUN

Segment cluster T08446_PEA_1_node_7 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 808 below describes the starting and ending position of this segment on each transcript.

Table 808- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	471	551
T08446_PEA_1_T22	471	551

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 809.

Table 809- Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
T08446_0_9_0	lung malignant tumors	LUN

Segment cluster T08446_PEA_1_node_12 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2 and T08446_PEA_1_T22. Table 810 below describes the starting and ending position of this segment on each transcript.

Table 810- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	690	782
T08446_PEA_1_T22	690	782

Segment cluster T08446_PEA_1_node_13 according to the present invention is supported by 0 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T22. Table 811 below describes the starting and ending position of this segment on each transcript.

Table 811 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T22	783	828

Segment cluster T08446_PEA_1_node_19 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 812 below describes the starting and ending position of this segment on each transcript.

Table 812 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	906	983

Segment cluster T08446_PEA_1_node_21 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 813 below describes the starting and ending position of this segment on each transcript.

Table 813 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	984	1050

Segment cluster T08446_PEA_1_node_23 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 814 below describes the starting and ending position of this segment on each transcript.

Table 814 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1051	1110

Segment cluster T08446_PEA_1_node_27 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 815 below describes the starting and ending position of this segment on each transcript.

Table 815 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

T08446_PEA_1_T2	1264	1366

Segment cluster T08446_PEA_1_node_32 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 816 below describes the starting and ending position of this segment on each transcript.

Table 816- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1512	1594

Segment cluster T08446_PEA_1_node_34 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 817 below describes the starting and ending position of this segment on each transcript.

Table 817- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	1595	1702

Segment cluster T08446_PEA_1_node_45 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 818 below describes the starting and ending position of this segment on each transcript.

Table 818- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2003	2091

Segment cluster T08446_PEA_1_node_46 according to the present invention is supported by 18 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 819 below describes the starting and ending position of this segment on each transcript.

Table 819- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2092	2148

Segment cluster T08446_PEA_1_node_48 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 820 below describes the starting and ending position of this segment on each transcript.

Table 820- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2149	2223

Segment cluster T08446_PEA_1_node_54 according to the present invention can be found in the following transcript(s): T08446_PEA_1_T2. Table 821 below describes the starting and ending position of this segment on each transcript.

Table 821- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	2695	2706

Segment cluster T08446_PEA_1_node_58 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 822 below describes the starting and ending position of this segment on each transcript.

Table 822- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3276	3359

Segment cluster T08446_PEA_1_node_60 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 823 below describes the starting and ending position of this segment on each transcript.

Table 823 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3671	3720

Segment cluster T08446_PEA_1_node_61 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 824 below describes the starting and ending position of this segment on each transcript.

Table 824 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	3721	3782

Segment cluster T08446_PEA_1_node_64 according to the present invention can be found in the following transcript(s): T08446_PEA_1_T2. Table 825 below describes the starting and ending position of this segment on each transcript.

Table 825 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	4415	4420

Segment cluster T08446_PEA_1_node_65 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 826 below describes the starting and ending position of this segment on each transcript.

Table 826 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	4421	4472

Segment cluster T08446_PEA_1_node_66 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T08446_PEA_1_T2. Table 827 below describes the starting and ending position of this segment on each transcript.

Table 827 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T08446_PEA_1_T2	4473	4539

Variant protein alignment to the previously known protein:

```
Sequence name: SNXQ_HUMAN

Sequence documentation:

Alignment of: T08446_PEA_1_P18 x SNXQ_HUMAN    ..

Alignment segment 1/1:

                      Quality: 1835.00
Escore:        0
              Matching length:      185                 Total
length:      185
 Matching Percent Similarity:   100.00   Matching Percent
Identity:   100.00
    Total Percent Similarity:   100.00      Total Percent
Identity:   100.00
                         Gaps:        0

Alignment:
```

```
                .         .         .         .         .
      1 MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKP 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKP 50
                .         .         .         .         .
     51 GKRLSAPRGPFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GKRLSAPRGPFPRLADCAHFHYENVDFGHIQLLLSPDREGPSLSGENELV 100
                .         .         .         .         .
    101 FGVQVTCQGRSWPVLRSYDDFRSLDAHLHRCIFDRRFSCLPELPPPPEGA 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 FGVQVTCQGRSWPVLRSYDDFRSLDAHLHRCIFDRRFSCLPELPPPPEGA 150
                .         .         .
    151 RAAQMLVPLLLQYLETLSGLVDSNLNCGPVLTWME               185
        |||||||||||||||||||||||||||||||||
    151 RAAQMLVPLLLQYLETLSGLVDSNLNCGPVLTWME               185
```

Sequence name: Q9NT23

Sequence documentation:

Alignment of: T08446_PEA_1_P18 x Q9NT23 ..

Alignment segment 1/1:

887

Quality: 8548.00

Escore: 0

Matching length: 862 Total length: 862

Matching Percent Similarity: 99.88 Matching Percent Identity: 99.88

Total Percent Similarity: 99.88 Total Percent Identity: 99.88

Gaps: 0

Alignment:

```
                .         .         .         .         .
    444 HDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRS 493
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      1 HDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRS 50
                .         .         .         .         .
    494 MELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRC 543
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     51 MELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRC 100
                .         .         .         .         .
    544 LLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERR 593
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    101 LLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERR 150
                .         .         .         .         .
    594 KGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSR 643
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    151 KGERGEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSR 200
                .         .         .         .         .
    644 PDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSC 693
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    201 PDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSC 250
```

```
694 ESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELD 743
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 ESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELD 300

744 FSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVT 793
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 FSPPRCLEGLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVT 350

794 PQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPA 843
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 PQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPA 400

844 LSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPL 893
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 LSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPL 450

894 PPPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPP 943
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 PPPPLSLLRPGGAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPP 500

944 ASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQS 993
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 ASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQS 550

994 DGSLLRSQRPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPS 1043
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 DGSLLRSQRPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPS 600

1044 QVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGP 1093
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 601 QVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGP 650
                .         .         .         .         .
1094 PAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQS 1143
     |||||||||||||||||||||||| ||||||||||||||||||||||||
 651 PAPLDRGENLYYEIGASEGSPYSGLTRSWSPFRSMPPDRLNASYGMLGQS 700
                .         .         .         .         .
1144 PPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALG 1193
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 701 PPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALG 750
                .         .         .         .         .
1194 PRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGP 1243
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 751 PRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGP 800
                .         .         .         .         .
1244 WGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSW 1293
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 801 WGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSW 850
                .
1294 SLHSEGQTRSYC                                       1305
     ||||||||||||
 851 SLHSEGQTRSYC                                       862
```

Sequence name: Q96CP3

Sequence documentation:

Alignment of: T08446_PEA_1_P18 x Q96CP3 ..

Alignment segment 1/1:

Quality: 3019.00

Escore: 0

Matching length: 295 Total length: 295

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
    1011 LRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLP 1060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 LRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLP 50
                .         .         .         .         .
    1061 PFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYEIGAS 1110
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 PFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYEIGAS 100
                .         .         .         .         .
    1111 EGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAP 1160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 EGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAP 150
                .         .         .         .         .
    1161 SCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAH 1210
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     151 SCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAH 200
```

```
                .         .         .         .         .
1211 PRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY 1260
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 201 PRSRSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAY 250
                .         .         .         .
1261 GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC      1305
     |||||||||||||||||||||||||||||||||||||||||||||
 251 GRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC      295
```

Sequence name: BAC86902

Sequence documentation:

Alignment of: T08446_PEA_1_P18 x BAC86902 ..

Alignment segment 1/1:

Quality: 9651.00

Escore: 0

Matching length: 991 Total length: 1019

Matching Percent Similarity: 99.90 Matching Percent Identity: 99.90

Total Percent Similarity: 97.15 Total Percent Identity: 97.15

Gaps: 1

Alignment:

```
                .         .         .         .         .
       155 MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIP 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
         1 MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIP 50
                .         .         .         .         .
       205 AVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSWWRGKRGFQVG 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
        51 AVAAAHVIKRYTAQAPDELSFEVGDIVSVIDMPPTEDRSWWRGKRGFQVG 100
                .         .         .         .         .
       255 FFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA 304
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       101 FFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLA 150
                .         .         .         .         .
       305 GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCS 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       151 GLLRTFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCS 200
                .         .         .         .         .
       355 EFIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVS 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       201 EFIEAHGVVDGIYRLSGVSSNIQRLRHEFDSERIPELSGPAFLQDIHSVS 250
                .         .         .         .         .
       405 SLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPH 454
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       251 SLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPH 300
                .         .         .         .         .
       455 YRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGA 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
       301 YRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGA 350
                .         .         .         .         .
```

505 AAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGS 554

351 AAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGS 400

555 CPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKP 604

401 CPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKP 450

605 GGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKS 654

451 GGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKS 500

655 EESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSES 704

501 EESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSCESLSSSSSSES 550

705 SSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLR 754

551 SSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLR 600

755 GLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRGPTS 804

601 GLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRGPTS 650

805 PASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHL 854

651 PASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPGRSLRPHL 700

855 IPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPG 904

701 IPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPG 750

```
905 GAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS 954
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 GAPPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLS 800

955 LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPM 1004
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 LEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQRPM 850

1005 GTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPA 1054
     ||||||||||||                                ||||||||||||
 851 GTSRRGLRGPA............................QVPTPGFFSPA 872

1055 PRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLY 1104
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 873 PRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLY 922

1105 YEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLL 1154
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 923 YEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLL 972

1155 SYPPAPSCFPPDHLGYSAP                                1173
     |||||||||||||||| |
 973 SYPPAPSCFPPDHLGYSPP                                991
```

DESCRIPTION FOR CLUSTER HUMCA1XIA

Cluster HUMCA1XIA features 4 transcript(s) and 46 segment(s) of interest, the names for which are given in Tables 828 and 829, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 830

Table 828 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMCA1XIA_T16	99
HUMCA1XIA_T17	100
HUMCA1XIA_T19	101
HUMCA1XIA_T20	102

Table 829 - Segments of interest

Segment Name	Sequence ID No.
HUMCA1XIA_node_0	742
HUMCA1XIA_node_2	743
HUMCA1XIA_node_4	744
HUMCA1XIA_node_6	745
HUMCA1XIA_node_8	746
HUMCA1XIA_node_9	747
HUMCA1XIA_node_18	748
HUMCA1XIA_node_54	749
HUMCA1XIA_node_55	750
HUMCA1XIA_node_92	751
HUMCA1XIA_node_11	752
HUMCA1XIA_node_15	753
HUMCA1XIA_node_19	754
HUMCA1XIA_node_21	755
HUMCA1XIA_node_23	756
HUMCA1XIA_node_25	757
HUMCA1XIA_node_27	758
HUMCA1XIA_node_29	759
HUMCA1XIA_node_31	760
HUMCA1XIA_node_33	761
HUMCA1XIA_node_35	762

HUMCA1XIA_node_37	763
HUMCA1XIA_node_39	764
HUMCA1XIA_node_41	765
HUMCA1XIA_node_43	766
HUMCA1XIA_node_45	767
HUMCA1XIA_node_47	769
HUMCA1XIA_node_49	769
HUMCA1XIA_node_51	770
HUMCA1XIA_node_57	771
HUMCA1XIA_node_59	772
HUMCA1XIA_node_62	773
HUMCA1XIA_node_64	774
HUMCA1XIA_node_66	775
HUMCA1XIA_node_68	776
HUMCA1XIA_node_70	777
HUMCA1XIA_node_72	778
HUMCA1XIA_node_74	779
HUMCA1XIA_node_76	780
HUMCA1XIA_node_78	782
HUMCA1XIA_node_81	783
HUMCA1XIA_node_83	784
HUMCA1XIA_node_85	785
HUMCA1XIA_node_87	786
HUMCA1XIA_node_89	787
HUMCA1XIA_node_91	788

Table 830 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HUMCA1XIA_P14	1372	HUMCA1XIA_T16

HUMCA1XIA_P15	1373	HUMCA1XIA_T17
HUMCA1XIA_P16	1374	HUMCA1XIA_T19
HUMCA1XIA_P17	1375	HUMCA1XIA_T20

These sequences are variants of the known protein Collagen alpha 1 (SwissProt accession identifier CA1B_HUMAN), SEQ ID NO: 1446, referred to herein as the previously known protein.

Protein Collagen alpha 1 is known or believed to have the following function(s): May play an important role in fibrillogenesis by controlling lateral growth of collagen II fibrils. The sequence for protein Collagen alpha 1 is given at the end of the application, as "Collagen alpha 1 amino acid sequence". Known polymorphisms for this sequence are as shown in Table 831.

Table 831 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
625	G -> V (in STL2). /FTId=VAR_013583.
676	G -> R (in STL2; overlapping phenotype with Marshall syndrome). /FTId=VAR_013584.
921 - 926	Missing (in STL2; overlapping phenotype with Marshall syndrome). /FTId=VAR_013585.
1313 - 1315	Missing (in STL2; overlapping phenotype with Marshall syndrome). /FTId=VAR_013586.
1516	G -> V (in STL2; overlapping phenotype with Marshall syndrome). /FTId=VAR_013587.
941 - 944	KDGL -> RMGC
986	Y -> H
1074	R -> P
1142	G -> D
1218	M -> W
1758	T -> A
1786	S -> N

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: cartilage condensation; vision; hearing; cell-cell adhesion; extracellular matrix organization and biogenesis, which are annotation(s) related to Biological Process; extracellular matrix structural protein; extracellular matrix protein, adhesive, which are annotation(s) related to Molecular Function; and extracellular matrix; collagen; collagen type XI, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HUMCA1XIA can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 32 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 32 and Table 832. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: bone malignant tumors, epithelial malignant tumors, a mixture of malignant tumors from different tissues and lung malignant tumors.

Table 832 - Normal tissue distribution

Name of Tissue	Number
adrenal	0
bone	207
brain	13
colon	0
epithelial	11
general	11

head and neck	0
kidney	0
lung	0
breast	8
pancreas	0
stomach	73
uterus	9

Table 833 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	4.2e-01	1.9e-01	9.6e-02	3.4	8.2e-02	3.6
bone	2.4e-01	6.3e-01	7.7e-10	4.3	5.3e-03	1.6
brain	5.0e-01	6.9e-01	1.8e-01	2.1	4.2e-01	1.3
colon	1.3e-02	2.9e-02	2.4e-01	3.0	3.5e-01	2.4
epithelial	3.9e-04	3.2e-03	1.3e-03	2.3	1.8e-02	1.7
general	5.6e-05	1.6e-03	9.5e-17	4.5	1.1e-09	2.8
head and neck	1.2e-01	2.1e-01	1	1.3	1	1.1
kidney	6.5e-01	7.2e-01	3.4e-01	2.4	4.9e-01	1.9
lung	5.3e-02	9.1e-02	5.5e-05	7.3	5.0e-03	4.0
breast	4.3e-01	5.6e-01	6.9e-01	1.4	8.2e-01	1.1
pancreas	3.3e-01	1.8e-01	4.2e-01	2.4	1.5e-01	3.7
stomach	5.0e-01	6.1e-01	6.9e-01	1.0	6.7e-01	0.8
Uterus	7.1e-01	7.0e-01	6.6e-01	1.1	6.4e-01	1.1

As noted above, cluster HUMCA1XIA features 4 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Collagen alpha 1. A description of each variant protein according to the present invention is now provided.

Variant protein HUMCA1XIA_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCA1XIA_T16. An alignment is given to the known protein (Collagen alpha 1) at the end

of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCA1XIA_P14 and CA1B_HUMAN_V5 (SEQ ID NO: 1447):

1.An isolated chimeric polypeptide encoding for HUMCA1XIA_P14, comprising a first amino acid sequence being at least 90 % homologous to
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT
GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY
NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM
IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH
YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT
EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED
TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN
GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG
RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM
GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP
GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAG
PRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQG
PIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPPGPQGPIGYPGPRGVK
GADGVRGLKGSKGEKGEDGFPGFKGDMGLKGDRGEVGQIGPRGEDGPEGPKGRAGPT
GDPGPSGQAGEKGKLGVPGLPGYPGRQGPKGSTGFPGFPGANGEKGARGVAGKPGPR
GQRGPTGPRGSRGARGPTGKPGPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGP
PGKDGLPGHPGQRGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQG
LPGAAGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQGPPGP
V corresponding to amino acids 1 - 1056 of CA1B_HUMAN_V5, which also corresponds to amino acids 1 - 1056 of HUMCA1XIA_P14, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence
VSMMIINSQTIMVVNYSSSFITLML corresponding to amino acids 1057 - 1081 of

HUMCA1XIA_P14, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMCA1XIA_P14, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSMMIINSQTIMVVNYSSSFITLML in HUMCA1XIA_P14.

It should be noted that the known protein sequence (CA1B_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for CA1B_HUMAN_V5. These changes were previously known to occur and are listed in the table below.

Table 834 - Changes to CA1B_HUMAN_V5

SNP position(s) on amino acid sequence	Type of change
987	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCA1XIA_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 835, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 835 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
8	W -> G	Yes
46	D -> E	Yes
559	G -> S	Yes
832	G -> *	Yes
986	H -> Y	Yes
1061	I -> M	Yes
1070	V -> A	Yes

Variant protein HUMCA1XIA_P14 is encoded by the following transcript(s): HUMCA1XIA_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCA1XIA_T16 is shown in bold; this coding portion starts at position 319 and ends at position 3561. The transcript also has the following SNPs as listed in Table 836 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 836 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
157	A -> G	No
241	T -> A	Yes
340	T -> G	Yes
456	T -> G	Yes
1993	G -> A	Yes
2812	G -> T	Yes
3274	C -> T	Yes
3282	C -> T	Yes

3501	A -> G	Yes
3527	T -> C	Yes

Variant protein HUMCA1XIA_P15 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCA1XIA_T17. An alignment is given to the known protein (Collagen alpha 1) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCA1XIA_P15 and CA1B_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCA1XIA_P15, comprising a first amino acid sequence being at least 90 % homologous to MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAG PRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQG PIGPPGEK corresponding to amino acids 1 - 714 of CA1B_HUMAN, which also corresponds to amino acids 1 - 714 of HUMCA1XIA_P15, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MCCNLSFGILIPLQK corresponding to amino acids 715 - 729 of HUMCA1XIA_P15, wherein

said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMCA1XIA_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MCCNLSFGILIPLQK in HUMCA1XIA_P15.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCA1XIA_P15 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 837, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 837- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
8	W -> G	Yes
46	D -> E	Yes
559	G -> S	Yes

The glycosylation sites of variant protein HUMCA1XIA_P15, as compared to the known protein Collagen alpha 1, are described in Table 838 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation

site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 838 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
1640	no

Variant protein HUMCA1XIA_P15 is encoded by the following transcript(s): HUMCA1XIA_T17, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCA1XIA_T17 is shown in bold; this coding portion starts at position 319 and ends at position 2505. The transcript also has the following SNPs as listed in Table 839 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 839 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
157	A -> G	No
241	T -> A	Yes
340	T -> G	Yes
456	T -> G	Yes
1993	G -> A	Yes
2473	C -> T	Yes

Variant protein HUMCA1XIA_P16 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCA1XIA_T19. An alignment is given to the known protein (Collagen alpha 1) at the end of the application. One or more alignments to one or more previously published protein

sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCA1XIA_P16 and CA1B_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCA1XIA_P16, comprising a first amino acid sequence being at least 90 % homologous to MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH YSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQT EANIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSED TLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEEFGPGVPAETDITETSIN GHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPTGPPGDPGDRGPPG RPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPM GLTGRPGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMP GEPGAKGDRGFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEA corresponding to amino acids 1 - 648 of CA1B_HUMAN, which also corresponds to amino acids 1 - 648 of HUMCA1XIA_P16, a second amino acid sequence being at least 90 % homologous to GMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQGPIGPPGEK corresponding to amino acids 667 - 714 of CA1B_HUMAN, which also corresponds to amino acids 649 - 696 of HUMCA1XIA_P16, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VSFSFSLFYKKVIKFACDKRFVGRHDERKVVKLSLPLYLIYE corresponding to amino acids 697 - 738 of HUMCA1XIA_P16, wherein said first amino acid sequence, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HUMCA1XIA_P16, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise AG, having a

structure as follows: a sequence starting from any of amino acid numbers 648-x to 648; and ending at any of amino acid numbers 649+ ((n-2) - x), in which x varies from 0 to n-2.

3.An isolated polypeptide encoding for a tail of HUMCA1XIA_P16, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSFSFSLFYKKVIKFACDKRFVGRHDERKVVKLSLPLYLIYE in HUMCA1XIA_P16.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCA1XIA_P16 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 840, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 840 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
8	W -> G	Yes
46	D -> E	Yes
559	G -> S	Yes

The glycosylation sites of variant protein HUMCA1XIA_P16, as compared to the known protein Collagen alpha 1, are described in Table 841 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation

site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 841 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
1640	no

Variant protein HUMCA1XIA_P16 is encoded by the following transcript(s): HUMCA1XIA_T19, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCA1XIA_T19 is shown in bold; this coding portion starts at position 319 and ends at position 2532. The transcript also has the following SNPs as listed in Table 842 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P16 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 842 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
157	A -> G	No
241	T -> A	Yes
340	T -> G	Yes
456	T -> G	Yes
1993	G -> A	Yes
2606	C -> A	Yes
2677	T -> G	Yes
2849	C -> T	Yes

Variant protein HUMCA1XIA_P17 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s)

HUMCA1XIA_T20. An alignment is given to the known protein (Collagen alpha 1) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCA1XIA_P17 and CA1B_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCA1XIA_P17, comprising a first amino acid sequence being at least 90 % homologous to MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTT GFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIY NEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLFRTVNIADGKWHRVAISVEKKTVTM IVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEH YSPDCDSSAPKAAQAQEPQIDE corresponding to amino acids 1 - 260 of CA1B_HUMAN, which also corresponds to amino acids 1 - 260 of HUMCA1XIA_P17, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRSTRPEKVFVFQ corresponding to amino acids 261 - 273 of HUMCA1XIA_P17, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMCA1XIA_P17, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRSTRPEKVFVFQ in HUMCA1XIA_P17.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCA1XIA_P17 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 843, (given according to their position(s) on the

amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 843 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
8	W -> G	Yes
46	D -> E	Yes

The glycosylation sites of variant protein HUMCA1XIA_P17, as compared to the known protein Collagen alpha 1, are described in Table 844 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 844 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
1640	no

Variant protein HUMCA1XIA_P17 is encoded by the following transcript(s): HUMCA1XIA_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCA1XIA_T20 is shown in bold; this coding portion starts at position 319 and ends at position 1137. The transcript also has the following SNPs as listed in Table 845 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCA1XIA_P17 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 845 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
157	A -> G	No
241	T -> A	Yes
340	T -> G	Yes
456	T -> G	Yes
1150	A -> C	Yes

As noted above, cluster HUMCA1XIA features 46 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMCA1XIA_node_0 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17, HUMCA1XIA_T19 and HUMCA1XIA_T20. Table 846 below describes the starting and ending position of this segment on each transcript.

Table 846 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1	424
HUMCA1XIA_T17	1	424
HUMCA1XIA_T19	1	424
HUMCA1XIA_T20	1	424

Segment cluster HUMCA1XIA_node_2 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17,

HUMCA1XIA_T19 and HUMCA1XIA_T20. Table 847 below describes the starting and ending position of this segment on each transcript.

Table 847 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	425	592
HUMCA1XIA_T17	425	592
HUMCA1XIA_T19	425	592
HUMCA1XIA_T20	425	592

Segment cluster HUMCA1XIA_node_4 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17, HUMCA1XIA_T19 and HUMCA1XIA_T20. Table 848 below describes the starting and ending position of this segment on each transcript.

Table 848 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	593	806
HUMCA1XIA_T17	593	806
HUMCA1XIA_T19	593	806
HUMCA1XIA_T20	593	806

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 849.

Table 849 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMCA1XIA_0_18_0	lung malignant tumors	LUN

Segment cluster HUMCA1XIA_node_6 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17, HUMCA1XIA_T19 and HUMCA1XIA_T20. Table 850 below describes the starting and ending position of this segment on each transcript.

Table 850 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	807	969
HUMCA1XIA_T17	807	969
HUMCA1XIA_T19	807	969
HUMCA1XIA_T20	807	969

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 851.

Table 851 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMCA1XIA_0_18_0	lung malignant tumors	LUN

Segment cluster HUMCA1XIA_node_8 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17,

HUMCA1XIA_T19 and HUMCA1XIA_T20. Table 852 below describes the starting and ending position of this segment on each transcript.

Table 852 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	970	1098
HUMCA1XIA_T17	970	1098
HUMCA1XIA_T19	970	1098
HUMCA1XIA_T20	970	1098

Segment cluster HUMCA1XIA_node_9 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T20. Table 853 below describes the starting and ending position of this segment on each transcript.

Table 853 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T20	1099	1271

Segment cluster HUMCA1XIA_node_18 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 854 below describes the starting and ending position of this segment on each transcript.

Table 854 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMCA1XIA_T16	1309	1522
HUMCA1XIA_T17	1309	1522
HUMCA1XIA_T19	1309	1522

Segment cluster HUMCA1XIA_node_54 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T19. Table 855 below describes the starting and ending position of this segment on each transcript.

Table 855 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T19	2407	2836

Segment cluster HUMCA1XIA_node_55 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 856 below describes the starting and ending position of this segment on each transcript.

Table 856 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T17	2461	2648
HUMCA1XIA_T19	2837	3475

Segment cluster HUMCA1XIA_node_92 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 857 below describes the starting and ending position of this segment on each transcript.

Table 857 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3487	3615

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMCA1XIA_node_11 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 858 below describes the starting and ending position of this segment on each transcript.

Table 858 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1099	1215
HUMCA1XIA_T17	1099	1215
HUMCA1XIA_T19	1099	1215

Segment cluster HUMCA1XIA_node_15 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 859 below describes the starting and ending position of this segment on each transcript.

Table 859 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMCA1XIA_T16	1216	1308
HUMCA1XIA_T17	1216	1308
HUMCA1XIA_T19	1216	1308

Segment cluster HUMCA1XIA_node_19 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 860 below describes the starting and ending position of this segment on each transcript.

Table 860 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1523	1563
HUMCA1XIA_T17	1523	1563
HUMCA1XIA_T19	1523	1563

Segment cluster HUMCA1XIA_node_21 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 861 below describes the starting and ending position of this segment on each transcript.

Table 861 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1564	1626
HUMCA1XIA_T17	1564	1626
HUMCA1XIA_T19	1564	1626

Segment cluster HUMCA1XIA_node_23 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 862 below describes the starting and ending position of this segment on each transcript.

Table 862 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1627	1668
HUMCA1XIA_T17	1627	1668
HUMCA1XIA_T19	1627	1668

Segment cluster HUMCA1XIA_node_25 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 863 below describes the starting and ending position of this segment on each transcript.

Table 863- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1669	1731
HUMCA1XIA_T17	1669	1731
HUMCA1XIA_T19	1669	1731

Segment cluster HUMCA1XIA_node_27 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 864 below describes the starting and ending position of this segment on each transcript.

Table 864 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1732	1806
HUMCA1XIA_T17	1732	1806
HUMCA1XIA_T19	1732	1806

Segment cluster HUMCA1XIA_node_29 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 865 below describes the starting and ending position of this segment on each transcript.

Table 865 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1807	1890
HUMCA1XIA_T17	1807	1890
HUMCA1XIA_T19	1807	1890

Segment cluster HUMCA1XIA_node_31 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 866 below describes the starting and ending position of this segment on each transcript.

Table 866- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1891	1947
HUMCA1XIA_T17	1891	1947
HUMCA1XIA_T19	1891	1947

Segment cluster HUMCA1XIA_node_33 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 867 below describes the starting and ending position of this segment on each transcript.

Table 867 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	1948	2001
HUMCA1XIA_T17	1948	2001
HUMCA1XIA_T19	1948	2001

Segment cluster HUMCA1XIA_node_35 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 868 below describes the starting and ending position of this segment on each transcript.

Table 868 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMCA1XIA_T16	2002	2055
HUMCA1XIA_T17	2002	2055
HUMCA1XIA_T19	2002	2055

Segment cluster HUMCA1XIA_node_37 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 869 below describes the starting and ending position of this segment on each transcript.

Table 869 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2056	2109
HUMCA1XIA_T17	2056	2109
HUMCA1XIA_T19	2056	2109

Segment cluster HUMCA1XIA_node_39 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 870 below describes the starting and ending position of this segment on each transcript.

Table 870 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2110	2163
HUMCA1XIA_T17	2110	2163
HUMCA1XIA_T19	2110	2163

Segment cluster HUMCA1XIA_node_41 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 871 below describes the starting and ending position of this segment on each transcript.

Table 871 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2164	2217
HUMCA1XIA_T17	2164	2217
HUMCA1XIA_T19	2164	2217

Segment cluster HUMCA1XIA_node_43 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 872 below describes the starting and ending position of this segment on each transcript.

Table 872 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2218	2262
HUMCA1XIA_T17	2218	2262
HUMCA1XIA_T19	2218	2262

Segment cluster HUMCA1XIA_node_45 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): HUMCA1XIA_T16 and HUMCA1XIA_T17. Table 873 below describes the starting and ending position of this segment on each transcript.

Table 873 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2263	2316
HUMCA1XIA_T17	2263	2316

Segment cluster HUMCA1XIA_node_47 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 874 below describes the starting and ending position of this segment on each transcript.

Table 874 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2317	2361
HUMCA1XIA_T17	2317	2361
HUMCA1XIA_T19	2263	2307

Segment cluster HUMCA1XIA_node_49 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 875 below describes the starting and ending position of this segment on each transcript.

Table 875 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2362	2415
HUMCA1XIA_T17	2362	2415
HUMCA1XIA_T19	2308	2361

Segment cluster HUMCA1XIA_node_51 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16, HUMCA1XIA_T17 and HUMCA1XIA_T19. Table 876 below describes the starting and ending position of this segment on each transcript.

Table 876 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2416	2460
HUMCA1XIA_T17	2416	2460
HUMCA1XIA_T19	2362	2406

Segment cluster HUMCA1XIA_node_57 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 877 below describes the starting and ending position of this segment on each transcript.

Table 877 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2461	2514

Segment cluster HUMCA1XIA_node_59 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 878 below describes the starting and ending position of this segment on each transcript.

Table 878 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2515	2559

Segment cluster HUMCA1XIA_node_62 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 879 below describes the starting and ending position of this segment on each transcript.

Table 879 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2560	2613

Segment cluster HUMCA1XIA_node_64 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 880 below describes the starting and ending position of this segment on each transcript.

Table 880 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2614	2658

Segment cluster HUMCA1XIA_node_66 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 881 below describes the starting and ending position of this segment on each transcript.

Table 881 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2659	2712

Segment cluster HUMCA1XIA_node_68 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 882 below describes the starting and ending position of this segment on each transcript.

Table 882 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2713	2820

Segment cluster HUMCA1XIA_node_70 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 883 below describes the starting and ending position of this segment on each transcript.

Table 883 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMCA1XIA_T16	2821	2874

Segment cluster HUMCA1XIA_node_72 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 884 below describes the starting and ending position of this segment on each transcript.

Table 884 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2875	2928

Segment cluster HUMCA1XIA_node_74 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 885 below describes the starting and ending position of this segment on each transcript.

Table 885 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2929	2973

Segment cluster HUMCA1XIA_node_76 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 886 below describes the starting and ending position of this segment on each transcript.

Table 886 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	2974	3027

Segment cluster HUMCA1XIA_node_78 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 887 below describes the starting and ending position of this segment on each transcript.

Table 887 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3028	3072

Segment cluster HUMCA1XIA_node_81 according to the present invention is supported by 8 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 888 below describes the starting and ending position of this segment on each transcript.

Table 888 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3073	3126

Segment cluster HUMCA1XIA_node_83 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 889 below describes the starting and ending position of this segment on each transcript.

Table 889 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3127	3180

Segment cluster HUMCA1XIA_node_85 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 890 below describes the starting and ending position of this segment on each transcript.

Table 890 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3181	3234

Segment cluster HUMCA1XIA_node_87 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 891 below describes the starting and ending position of this segment on each transcript.

Table 891 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3235	3342

Segment cluster HUMCA1XIA_node_89 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): HUMCA1XIA_T16. Table 892 below describes the starting and ending position of this segment on each transcript.

Table 892 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3343	3432

Segment cluster HUMCA1XIA_node_91 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCA1XIA_T16. Table 893 below describes the starting and ending position of this segment on each transcript.

Table 893 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCA1XIA_T16	3433	3486

Variant protein alignment to the previously known protein:

Sequence name: CA1B_HUMAN_V5

Sequence documentation:

Alignment of: HUMCA1XIA_P14 x CA1B_HUMAN_V5 ..

Alignment segment 1/1:

Quality: 10456.00

Escore: 0

Matching length: 1058 Total length: 1058

Matching Percent Similarity: 99.91 Matching Percent Identity: 99.91

Total Percent Similarity: 99.91 Total Percent Identity: 99.91

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
                .         .         .         .         .
 51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
                .         .         .         .         .
101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
                .         .         .         .         .
151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
                .         .         .         .         .
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
```

```
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
```

601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGP 650

601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGP 650

651 RGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 700

651 RGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 700

701 LPGPQGPIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPP 750

701 LPGPQGPIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPP 750

751 GPQGPIGYPGPRGVKGADGVRGLKGSKGEKGEDGFPGFKGDMGLKGDRGE 800

751 GPQGPIGYPGPRGVKGADGVRGLKGSKGEKGEDGFPGFKGDMGLKGDRGE 800

801 VGQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLPGYPGRQG 850

801 VGQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLPGYPGRQG 850

851 PKGSTGFPGFPGANGEKGARGVAGKPGPRGQRGPTGPRGSRGARGPTGKP 900

851 PKGSTGFPGFPGANGEKGARGVAGKPGPRGQRGPTGPRGSRGARGPTGKP 900

901 GPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQ 950

901 GPKGTSGGDGPPGPPGERGPQGPQGPVGFPGPKGPPGPPGKDGLPGHPGQ 950

951 RGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPG 1000

951 RGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPG 1000

```
                .         .         .         .         .
    1001 AAGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQ 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQ 1050

    1051 GPPGPVVS                                          1058
         |||||| |
    1051 GPPGPVGS                                          1058
```

Sequence name: CA1B_HUMAN

Sequence documentation:

Alignment of: HUMCA1XIA_P15 x CA1B_HUMAN ..

Alignment segment 1/1:

Quality: 7073.00

Escore: 0

Matching length: 714 Total length: 714

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                 .         .         .         .         .
           1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
             ||||||||||||||||||||||||||||||||||||||||||||||||||
           1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
                 .         .         .         .         .
          51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||
          51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
                 .         .         .         .         .
         101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
                 .         .         .         .         .
         151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
                 .         .         .         .         .
         201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
                 .         .         .         .         .
         251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
                 .         .         .         .         .
         301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
             ||||||||||||||||||||||||||||||||||||||||||||||||||
         301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
                 .         .         .         .         .
```

```
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
              .         .         .         .         .
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
              .         .         .         .         .
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
              .         .         .         .         .
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
              .         .         .         .         .
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
              .         .         .         .         .
601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGP 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGP 650
              .         .         .         .         .
651 RGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 700
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
651 RGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 700
              .
701 LPGPQGPIGPPGEK                                     714
    ||||||||||||||
701 LPGPQGPIGPPGEK                                     714
```

Sequence name: CA1B_HUMAN

Sequence documentation:

Alignment of: HUMCA1XIA_P16 x CA1B_HUMAN ..

Alignment segment 1/1:

Quality: 6795.00
Escore: 0
Matching length: 696 Total length: 714
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 97.48 Total Percent Identity: 97.48
Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
                .         .         .         .         .
       51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
```

```
                                                                                
                                                                                
 51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
                                                                                
101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
                                                                                
101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
                                                                                
151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
                                                                                
151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
                                                                                
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
                                                                                
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
                                                                                
251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
                                                                                
251 AQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEA 300
                                                                                
301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
                                                                                
301 NIVDDFQEYNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDS 350
                                                                                
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
                                                                                
351 QRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYEYKEYEDKPTSPPNEE 400
                                                                                
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
                                                                                
401 FGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPA 450
                                                                                
```

```
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 GIMGPPGLQGPTGPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYG 500
              .         .         .         .         .
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 GDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGRPGPVGGPGSSG 550
              .         .         .         .         .
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 AKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDR 600
              .         .         .         .         .
601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEA.. 648
    ||||||||||||||||||||||||||||||||||||||||||||||||
601 GFDGLPGLPGDKGHRGERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGP 650
              .         .         .         .         .
649 ................GMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 682
                    ||||||||||||||||||||||||||||||||
651 RGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQG 700
              .
683 LPGPQGPIGPPGEK                                     696
    ||||||||||||||
701 LPGPQGPIGPPGEK                                     714
```

Sequence name: CA1B_HUMAN

Sequence documentation:

Alignment of: HUMCA1XIA_P17 x CA1B_HUMAN ..

Alignment segment 1/1:

Quality: 2561.00

Escore: 0

Matching length: 260 Total length: 260

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNS 50
                .         .         .         .         .
     51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 PEGISKTTGFCTNRKNSKGSDTAYRVSKQAQLSAPTKQLFPGGTFPEDFS 100
                .         .         .         .         .
    101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPED 150
                .         .         .         .         .
    151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
```

941

```
                                                                    
151 YPLFRTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDT 200
                                                                    
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
                                                                    
201 NGITVFGTRILDEEVFEGDIQQFLITGDPKAAYDYCEHYSPDCDSSAPKA 250
                                                                    
251 AQAQEPQIDE                                                 260
                                                                    
251 AQAQEPQIDE                                                 260
```

Expression of Homo sapiens collagen, type XI, alpha 1 (COL11A1) HUMCA1X1A transcripts which are detectable by amplicon as depicted in sequence name HUMCA1X1A seg55 in normal and cancerous lung tissues

Expression of Homo sapiens collagen, type XI, alpha 1 (COL11A1) transcripts detectable by or according to seg55, HUMCA1X1A seg55 amplicon (SEQ ID NO:1663) and primers HUMCA1X1A seg55F (SEQ ID NO:1661) and HUMCA1X1A seg55R (SEQ ID NO:1662) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 67 is a histogram showing over expression of the above-indicated Homo sapiens collagen, type XI, alpha 1 (COL11A1) transcripts in cancerous lung samples relative to the

normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.

As is evident from Figure 67, the expression of Homo sapiens collagen, type XI, alpha 1 (COL11A1) transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 5 fold was found in 11 out of 15 adenocarcinoma samples, 11 out of 16 squamous cell carcinoma samples, and in 2 out of 4 large cell carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: HUMCA1X1A seg55F forward primer; and HUMCA1X1A seg55R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: HUMCA1X1A seg55.

Forward primer -HUMCA1X1A seg55F (SEQ ID NO:1661):
TTCTCATAGTATTCCATTGATTGGGTA
Reverse primer- HUMCA1X1A seg55R (SEQ ID NO:1662):
CACCGGTATGGAGAATAGCGA
Amplicon (SEQ ID NO:1663):
TTCTCATAGTATTCCATTGATTGGGTATACCAGGTTCTGTTTACTTTTACTTGGCAGT
TGATAGAATAGGTGTAGTTTATACTTTTTCGCTATTCTCCATACCGGTG

DESCRIPTION FOR CLUSTER T11628

Cluster T11628 features 6 transcript(s) and 25 segment(s) of interest, the names for which are given in Tables 894 and 895, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 896.

Table 894 - Transcripts of interest

Transcript Name	Sequence ID No.
T11628_PEA_1_T3	103
T11628_PEA_1_T4	104
T11628_PEA_1_T5	105
T11628_PEA_1_T7	106
T11628_PEA_1_T9	107
T11628_PEA_1_T11	108

Table 895 - Segments of interest

Segment Name	Sequence ID No.
T11628_PEA_1_node_7	789
T11628_PEA_1_node_11	790
T11628_PEA_1_node_16	791
T11628_PEA_1_node_22	792
T11628_PEA_1_node_25	793
T11628_PEA_1_node_31	794
T11628_PEA_1_node_37	795
T11628_PEA_1_node_0	796
T11628_PEA_1_node_4	797
T11628_PEA_1_node_9	798
T11628_PEA_1_node_13	799
T11628_PEA_1_node_14	800
T11628_PEA_1_node_17	801
T11628_PEA_1_node_18	802

T11628_PEA_1_node_19	803
T11628_PEA_1_node_24	804
T11628_PEA_1_node_27	805
T11628_PEA_1_node_28	806
T11628_PEA_1_node_29	807
T11628_PEA_1_node_30	808
T11628_PEA_1_node_32	809
T11628_PEA_1_node_33	810
T11628_PEA_1_node_34	811
T11628_PEA_1_node_35	812
T11628_PEA_1_node_36	813

Table 896 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
T11628_PEA_1_P2	1376	T11628_PEA_1_T3; T11628_PEA_1_T5; T11628_PEA_1_T7
T11628_PEA_1_P5	1377	T11628_PEA_1_T9
T11628_PEA_1_P7	1378	T11628_PEA_1_T11
T11628_PEA_1_P10	1379	T11628_PEA_1_T4

These sequences are variants of the known protein Myoglobin (SwissProt accession identifier MYG_HUMAN), SEQ ID NO: 1448, referred to herein as the previously known protein.

Protein Myoglobin is known or believed to have the following function(s): Serves as a reserve supply of oxygen and facilitates the movement of oxygen within muscles. The sequence for protein Myoglobin is given at the end of the application, as "Myoglobin amino acid sequence". Known polymorphisms for this sequence are as shown in Table 897.

Table 897 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
54	E -> K. /FTId=VAR_003180.
133	K -> N. /FTId=VAR_003181.
139	R -> Q. /FTId=VAR_003182.
139	R -> W. /FTId=VAR_003183.
128	Q -> E

As noted above, cluster T11628 features 6 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Myoglobin. A description of each variant protein according to the present invention is now provided.

Variant protein T11628_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T11628_PEA_1_T3. An alignment is given to the known protein (Myoglobin) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T11628_PEA_1_P2 and Q8WVH6 (SEQ ID NO:1450):

1.An isolated chimeric polypeptide encoding for T11628_PEA_1_P2, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE corresponding to amino acids 1 - 55 of T11628_PEA_1_P2, and a second amino acid sequence being at least 90 % homologous to MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 1 - 99 of Q8WVH6, which also corresponds to amino acids 56 - 154 of T11628_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of T11628_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE of T11628_PEA_1_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein T11628_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 898, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 898 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
26	G ->	No
44	F ->	No
92	Q -> R	No
135	A ->	No
141	K ->	No
153	Q ->	No

Variant protein T11628_PEA_1_P2 is encoded by the following transcript(s): T11628_PEA_1_T3, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T11628_PEA_1_T3 is shown in bold; this coding portion starts at position 220 and ends at position 681. The transcript also has the following SNPs as listed in Table 899 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 899- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
83	G -> A	Yes
93	G -> A	Yes
95	G -> A	Yes
146	G -> A	Yes
295	G ->	No
349	T ->	No
393	G -> A	Yes
423	C -> T	Yes
494	A -> G	No
498	G -> A	No
623	C ->	No
642	G ->	No
678	G ->	No
686	C ->	No
686	C -> A	No
717	C ->	No
787	T -> G	No
820	G -> T	No
826	G -> T	No

850	C ->	No
934	T -> G	No
975	A -> G	Yes
1117	G ->	No
1218	A -> G	No

Variant protein T11628_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T11628_PEA_1_T9. An alignment is given to the known protein (Myoglobin) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T11628_PEA_1_P5 and MYG_HUMAN_V1 (SEQ ID NO:1449):

1.An isolated chimeric polypeptide encoding for T11628_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 56 - 154 of MYG_HUMAN_V1, which also corresponds to amino acids 1 - 99 of T11628_PEA_1_P5.

It should be noted that the known protein sequence (MYG_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for MYG_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 900 - Changes to MYG_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein T11628_PEA_1_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 901, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 901 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
37	Q -> R	No
80	A ->	No
86	K ->	No
98	Q ->	No

Variant protein T11628_PEA_1_P5 is encoded by the following transcript(s): T11628_PEA_1_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T11628_PEA_1_T9 is shown in bold; this coding portion starts at position 211 and ends at position 507. The transcript also has the following SNPs as listed in Table 902 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein T11628_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 902 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
2	C -> T	Yes
175	T ->	No
219	G -> A	Yes
249	C -> T	Yes
320	A -> G	No
324	G -> A	No
449	C ->	No
468	G ->	No
504	G ->	No
512	C ->	No
512	C -> A	No
543	C ->	No
613	T -> G	No
646	G -> T	No
652	G -> T	No
676	C ->	No
760	T -> G	No
801	A -> G	Yes
943	G ->	No
1044	A -> G	No

Variant protein T11628_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T11628_PEA_1_T11. An alignment is given to the known protein (Myoglobin) at the end of the

application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T11628_PEA_1_P7 and MYG_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for T11628_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMK ASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQ SKHPGDFGADAQGAMNK corresponding to amino acids 1 - 134 of MYG_HUMAN_V1, which also corresponds to amino acids 1 - 134 of T11628_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence G corresponding to amino acids 135 - 135 of T11628_PEA_1_P7, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

It should be noted that the known protein sequence (MYG_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for MYG_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 903 - Changes to MYG_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein.

In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein T11628_PEA_1_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 904, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 904 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
26	G ->	No
44	F ->	No
92	Q -> R	No

Variant protein T11628_PEA_1_P7 is encoded by the following transcript(s): T11628_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T11628_PEA_1_T11 is shown in bold; this coding portion starts at position 319 and ends at position 723. The transcript also has the following SNPs as listed in Table 905 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 905 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
394	G ->	No
448	T ->	No
492	G -> A	Yes

522	C -> T	Yes
593	A -> G	No
597	G -> A	No
728	C ->	No
728	C -> A	No
759	C ->	No
829	T -> G	No
862	G -> T	No
868	G -> T	No
892	C ->	No
976	T -> G	No
1017	A -> G	Yes
1159	G ->	No
1260	A -> G	No

Variant protein T11628_PEA_1_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) T11628_PEA_1_T4. An alignment is given to the known protein (Myoglobin) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between T11628_PEA_1_P10 and Q8WVH6 (SEQ ID NO: 1450):

1.An isolated chimeric polypeptide encoding for T11628_PEA_1_P10, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE corresponding to amino acids 1 - 55 of T11628_PEA_1_P10, and a second amino acid sequence being at least 90 % homologous to MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQV

LQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG corresponding to amino acids 1 - 99 of Q8WVH6, which also corresponds to amino acids 56 - 154 of T11628_PEA_1_P10, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of T11628_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDE of T11628_PEA_1_P10.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein T11628_PEA_1_P10 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 906, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 906 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
26	G ->	No
44	F ->	No
92	Q -> R	No

135	A ->	No
141	K ->	No
153	Q ->	No

Variant protein T11628_PEA_1_P10 is encoded by the following transcript(s): T11628_PEA_1_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript T11628_PEA_1_T4 is shown in bold; this coding portion starts at position 205 and ends at position 666. The transcript also has the following SNPs as listed in Table 907 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein T11628_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 907 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
280	G ->	No
334	T ->	No
378	G -> A	Yes
408	C -> T	Yes
479	A -> G	No
483	G -> A	No
608	C ->	No
627	G ->	No
663	G ->	No
671	C ->	No
671	C -> A	No
702	C ->	No
772	T -> G	No
805	G -> T	No
811	G -> T	No

835	C ->	No
919	T -> G	No
960	A -> G	Yes
1102	G ->	No
1203	A -> G	No

As noted above, cluster T11628 features 25 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster T11628_PEA_1_node_7 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3. Table 908 below describes the starting and ending position of this segment on each transcript.

Table 908 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	1	211

Segment cluster T11628_PEA_1_node_11 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T5. Table 909 below describes the starting and ending position of this segment on each transcript.

Table 909 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T5	48	178

Segment cluster T11628_PEA_1_node_16 according to the present invention is supported by 38 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T11. Table 910 below describes the starting and ending position of this segment on each transcript.

Table 910 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T11	1	214

Segment cluster T11628_PEA_1_node_22 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T9. Table 911 below describes the starting and ending position of this segment on each transcript.

Table 911 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T9	1	140

Segment cluster T11628_PEA_1_node_25 according to the present invention is supported by 129 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 912 below describes the starting and ending position of this segment on each transcript.

Table 912- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

T11628_PEA_1_T3	395	537
T11628_PEA_1_T4	380	522
T11628_PEA_1_T5	362	504
T11628_PEA_1_T7	347	489
T11628_PEA_1_T9	221	363
T11628_PEA_1_T11	494	636

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 913.

Table 913 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
T11628_0_9_0	lung malignant tumors	LUN

Segment cluster T11628_PEA_1_node_31 according to the present invention is supported by 137 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 914 below describes the starting and ending position of this segment on each transcript.

Table 914 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	702	831
T11628_PEA_1_T4	687	816
T11628_PEA_1_T5	669	798
T11628_PEA_1_T7	654	783
T11628_PEA_1_T9	528	657

T11628_PEA_1_T11	744	873

Segment cluster T11628_PEA_1_node_37 according to the present invention is supported by 99 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 915 below describes the starting and ending position of this segment on each transcript.

Table 915 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	1086	1225
T11628_PEA_1_T4	1071	1210
T11628_PEA_1_T5	1053	1192
T11628_PEA_1_T7	1038	1177
T11628_PEA_1_T9	912	1051
T11628_PEA_1_T11	1128	1267

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster T11628_PEA_1_node_0 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T4. Table 916 below describes the starting and ending position of this segment on each transcript.

Table 916 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T4	1	93

Segment cluster T11628_PEA_1_node_4 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T4. Table 917 below describes the starting and ending position of this segment on each transcript.

Table 917 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T4	94	196

Segment cluster T11628_PEA_1_node_9 according to the present invention is supported by 16 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T5 and T11628_PEA_1_T7. Table 918 below describes the starting and ending position of this segment on each transcript.

Table 918 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T5	1	47
T11628_PEA_1_T7	1	47

Segment cluster T11628_PEA_1_node_13 according to the present invention can be found in the following transcript(s): T11628_PEA_1_T7. Table 919 below describes the starting and ending position of this segment on each transcript.

Table 919 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T7	48	65

Segment cluster T11628_PEA_1_node_14 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T7. Table 920 below describes the starting and ending position of this segment on each transcript.

Table 920 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T7	66	163

Segment cluster T11628_PEA_1_node_17 according to the present invention is supported by 55 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T11. Table 921 below describes the starting and ending position of this segment on each transcript.

Table 921 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T11	215	310

Segment cluster T11628_PEA_1_node_18 according to the present invention is supported by 98 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7 and T11628_PEA_1_T11. Table 922 below describes the starting and ending position of this segment on each transcript.

Table 922 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	212	289
T11628_PEA_1_T4	197	274
T11628_PEA_1_T5	179	256
T11628_PEA_1_T7	164	241
T11628_PEA_1_T11	311	388

Segment cluster T11628_PEA_1_node_19 according to the present invention can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7 and T11628_PEA_1_T11. Table 923 below describes the starting and ending position of this segment on each transcript.

Table 923 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	290	314
T11628_PEA_1_T4	275	299
T11628_PEA_1_T5	257	281
T11628_PEA_1_T7	242	266
T11628_PEA_1_T11	389	413

Segment cluster T11628_PEA_1_node_24 according to the present invention is supported by 112 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 924 below describes the starting and ending position of this segment on each transcript.

Table 924 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	315	394
T11628_PEA_1_T4	300	379
T11628_PEA_1_T5	282	361
T11628_PEA_1_T7	267	346
T11628_PEA_1_T9	141	220
T11628_PEA_1_T11	414	493

Segment cluster T11628_PEA_1_node_27 according to the present invention is supported by 119 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 925 below describes the starting and ending position of this segment on each transcript.

Table 925 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	538	621
T11628_PEA_1_T4	523	606
T11628_PEA_1_T5	505	588
T11628_PEA_1_T7	490	573
T11628_PEA_1_T9	364	447
T11628_PEA_1_T11	637	720

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 926

Table 926 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
T11628_0_9_0	lung malignant tumors	LUN

Segment cluster T11628_PEA_1_node_28 according to the present invention is supported by 115 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7 and T11628_PEA_1_T9. Table 927 below describes the starting and ending position of this segment on each transcript.

Table 927 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	622	650
T11628_PEA_1_T4	607	635
T11628_PEA_1_T5	589	617
T11628_PEA_1_T7	574	602
T11628_PEA_1_T9	448	476

Segment cluster T11628_PEA_1_node_29 according to the present invention is supported by 113 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7 and T11628_PEA_1_T9. Table 928 below describes the starting and ending position of this segment on each transcript.

Table 928 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	651	678

T11628_PEA_1_T4	636	663
T11628_PEA_1_T5	618	645
T11628_PEA_1_T7	603	630
T11628_PEA_1_T9	477	504

Segment cluster T11628_PEA_1_node_30 according to the present invention can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 929 below describes the starting and ending position of this segment on each transcript.

Table 929 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	679	701
T11628_PEA_1_T4	664	686
T11628_PEA_1_T5	646	668
T11628_PEA_1_T7	631	653
T11628_PEA_1_T9	505	527
T11628_PEA_1_T11	721	743

Segment cluster T11628_PEA_1_node_32 according to the present invention can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 930 below describes the starting and ending position of this segment on each transcript.

Table 930 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	832	844

T11628_PEA_1_T4	817	829
T11628_PEA_1_T5	799	811
T11628_PEA_1_T7	784	796
T11628_PEA_1_T9	658	670
T11628_PEA_1_T11	874	886

Segment cluster T11628_PEA_1_node_33 according to the present invention can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 931 below describes the starting and ending position of this segment on each transcript.

Table 931 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	845	866
T11628_PEA_1_T4	830	851
T11628_PEA_1_T5	812	833
T11628_PEA_1_T7	797	818
T11628_PEA_1_T9	671	692
T11628_PEA_1_T11	887	908

Segment cluster T11628_PEA_1_node_34 according to the present invention is supported by 122 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 932 below describes the starting and ending position of this segment on each transcript.

Table 932 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	867	911
T11628_PEA_1_T4	852	896
T11628_PEA_1_T5	834	878
T11628_PEA_1_T7	819	863
T11628_PEA_1_T9	693	737
T11628_PEA_1_T11	909	953

Segment cluster T11628_PEA_1_node_35 according to the present invention is supported by 126 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4, T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 933 below describes the starting and ending position of this segment on each transcript.

Table 933 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	912	967
T11628_PEA_1_T4	897	952
T11628_PEA_1_T5	879	934
T11628_PEA_1_T7	864	919
T11628_PEA_1_T9	738	793
T11628_PEA_1_T11	954	1009

Segment cluster T11628_PEA_1_node_36 according to the present invention is supported by 122 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): T11628_PEA_1_T3, T11628_PEA_1_T4,

T11628_PEA_1_T5, T11628_PEA_1_T7, T11628_PEA_1_T9 and T11628_PEA_1_T11. Table 934 below describes the starting and ending position of this segment on each transcript.

Table 934 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
T11628_PEA_1_T3	968	1085
T11628_PEA_1_T4	953	1070
T11628_PEA_1_T5	935	1052
T11628_PEA_1_T7	920	1037
T11628_PEA_1_T9	794	911
T11628_PEA_1_T11	1010	1127

Variant protein alignment to the previously known protein:

```
Sequence name: Q8WVH6

Sequence documentation:

Alignment of: T11628_PEA_1_P2 x Q8WVH6    ..

Alignment segment 1/1:

                        Quality:  962.00
Escore:        0
               Matching length:       99                  Total
length:       99
```

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      56 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 50
                .         .         .         .
     106 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG 99
```

Sequence name: MYG_HUMAN_V1

Sequence documentation:

Alignment of: T11628_PEA_1_P5 x MYG_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 962.00

Escore: 0

970

Matching length: 99 Total length: 99

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                    .         .         .         .         .
        1 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       56 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 105
                    .         .         .         .
       51 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG  99
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      106 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG 154
```

Sequence name: MYG_HUMAN_V1

Sequence documentation:

Alignment of: T11628_PEA_1_P7 x MYG_HUMAN_V1 ..

Alignment segment 1/1:

971

Quality: 1315.00

Escore: 0

Matching length: 134 Total length: 134

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHL 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHL 50
                .         .         .         .         .
       51 KSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 KSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKI 100
                .         .         .
      101 PVKYLEFISECIIQVLQSKHPGDFGADAQGAMNK                 134
          ||||||||||||||||||||||||||||||||
      101 PVKYLEFISECIIQVLQSKHPGDFGADAQGAMNK                 134
```

Sequence name: Q8WVH6

Sequence documentation:

Alignment of: T11628_PEA_1_P10 x Q8WVH6 ..

Alignment segment 1/1:

Quality: 962.00

Escore: 0

Matching length: 99 Total length: 99

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      56 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 105
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYL 50
                .         .         .         .
     106 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG  154
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 EFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG  99
```

DESCRIPTION FOR CLUSTER HUMCEA

Cluster HUMCEA features 5 transcript(s) and 42 segment(s) of interest, the names for which are given in Tables 935 and 936, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 937.

Table 935 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMCEA_PEA_1_T8	109
HUMCEA_PEA_1_T9	110
HUMCEA_PEA_1_T20	111
HUMCEA_PEA_1_T25	112
HUMCEA_PEA_1_T26	113

Table 936 - Segments of interest

Segment Name	Sequence ID No.
HUMCEA_PEA_1_node_0	814
HUMCEA_PEA_1_node_2	815
HUMCEA_PEA_1_node_11	816
HUMCEA_PEA_1_node_12	817
HUMCEA_PEA_1_node_31	818
HUMCEA_PEA_1_node_36	819
HUMCEA_PEA_1_node_44	820
HUMCEA_PEA_1_node_46	821
HUMCEA_PEA_1_node_63	822
HUMCEA_PEA_1_node_65	823
HUMCEA_PEA_1_node_67	824
HUMCEA_PEA_1_node_3	825
HUMCEA_PEA_1_node_7	826
HUMCEA_PEA_1_node_8	827
HUMCEA_PEA_1_node_9	828
HUMCEA_PEA_1_node_10	829
HUMCEA_PEA_1_node_15	830
HUMCEA_PEA_1_node_16	831
HUMCEA_PEA_1_node_17	832

HUMCEA_PEA_1_node_18	833
HUMCEA_PEA_1_node_19	834
HUMCEA_PEA_1_node_20	835
HUMCEA_PEA_1_node_21	836
HUMCEA_PEA_1_node_22	837
HUMCEA_PEA_1_node_23	838
HUMCEA_PEA_1_node_24	839
HUMCEA_PEA_1_node_27	840
HUMCEA_PEA_1_node_29	841
HUMCEA_PEA_1_node_30	842
HUMCEA_PEA_1_node_33	843
HUMCEA_PEA_1_node_34	844
HUMCEA_PEA_1_node_35	845
HUMCEA_PEA_1_node_45	846
HUMCEA_PEA_1_node_50	847
HUMCEA_PEA_1_node_51	848
HUMCEA_PEA_1_node_56	849
HUMCEA_PEA_1_node_57	850
HUMCEA_PEA_1_node_58	851
HUMCEA_PEA_1_node_60	852
HUMCEA_PEA_1_node_61	853
HUMCEA_PEA_1_node_62	854
HUMCEA_PEA_1_node_64	855

Table 937 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HUMCEA_PEA_1_P4	1380	HUMCEA_PEA_1_T8
HUMCEA_PEA_1_P5	1381	HUMCEA_PEA_1_T9
HUMCEA_PEA_1_P14	1382	HUMCEA_PEA_1_T20

HUMCEA_PEA_1_P19	1383	HUMCEA_PEA_1_T25
HUMCEA_PEA_1_P20	1384	HUMCEA_PEA_1_T26

These sequences are variants of the known protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor (SwissProt accession identifier CEA5_HUMAN; known also according to the synonyms Carcinoembryonic antigen; CEA; Meconium antigen 100; CD66e antigen), SEQ ID NO: 1451, referred to herein as the previously known protein.

The sequence for protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor is given at the end of the application, as "Carcinoembryonic antigen-related cell adhesion molecule 5 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 938

Table 938 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
320	Missing

Protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor localization is believed to be attached to the membrane by a GPI-anchor.

The previously known protein also has the following indication(s) and/or potential therapeutic use(s): Cancer. It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Immunostimulant. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Imaging agent; Anticancer; Immunostimulant; Immunoconjugate; Monoclonal antibody, murine; Antisense therapy; antibody.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: integral plasma membrane protein; membrane, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HUMCEA can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 33 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 33 and Table 939. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Table 939 - Normal tissue distribution

Name of Tissue	Number
colon	1175
epithelial	92
general	29
head and neck	81
kidney	0
lung	0
lymph nodes	0
breast	0
pancreas	0

prostate	0
stomach	256

Table 940 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
colon	2.0e-01	2.7e-01	9.8e-01	0.5	1	0.5
epithelial	2.1e-03	2.7e-02	6.4e-04	1.4	2.1e-01	1.0
general	3.9e-08	8.2e-06	9.2e-18	3.2	1.3e-10	2.2
head and neck	3.4e-01	5.0e-01	2.1e-01	1.8	5.6e-01	0.9
kidney	4.3e-01	5.3e-01	5.8e-01	2.1	7.0e-01	1.6
lung	1.3e-01	2.6e-01	1	1.1	1	1.1
lymph nodes	3.1e-01	5.7e-01	8.1e-02	6.0	3.3e-01	2.5
breast	3.8e-01	1.5e-01	1	1.0	6.8e-01	1.5
pancreas	2.2e-02	2.3e-02	1.4e-08	7.8	7.4e-07	6.4
prostate	5.3e-01	6.0e-01	3.0e-01	2.5	4.2e-01	2.0
stomach	1.5e-01	4.7e-01	8.9e-01	0.6	7.2e-01	0.4

For this cluster, at least one oligonucleotide was found to demonstrate overexpression of the cluster, although not of at least one transcript/segment as listed below. Microarray (chip) data is also available for this cluster as follows. Various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer, as previously described. The following oligonucleotides were found to hit this cluster but not other segments/transcripts below (in relation to lung cancer), shown in Table 941.

Table 941 - Oligonucleotides related to this cluster

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMCEA_0_0_15168	lung malignant tumors	LUN

As noted above, cluster HUMCEA features 5 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HUMCEA_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCEA_PEA_1_T8. An alignment is given to the known protein (Carcinoembryonic antigen-related cell adhesion molecule 5 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCEA_PEA_1_P4 and CEA5_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWV NNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVL corresponding to amino acids 1 - 234 of CEA5_HUMAN, which also corresponds to amino acids 1 - 234 of HUMCEA_PEA_1_P4, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence CEYICSSLAQAASPNPQGQRQDFSVPLRFKYTDPQPWTSRLSVTFCPRKTWADQVLTKN RRGGAASVLGGSGSTPYDGRNR corresponding to amino acids 235 - 315 of HUMCEA_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMCEA_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence

CEYICSSLAQAASPNPQGQRQDFSVPLRFKYTDPQPWTSRLSVTFCPRKTWADQVLTKN RRGGAASVLGGSGSTPYDGRNR in HUMCEA_PEA_1_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCEA_PEA_1_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 942, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 942 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
63	F -> L	No
80	I -> V	Yes
83	V -> A	Yes
137	Q -> P	Yes
173	D -> N	No

The glycosylation sites of variant protein HUMCEA_PEA_1_P4, as compared to the known protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor, are described in Table 943 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 943 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
197	yes	197
466	no	
360	no	
288	no	
665	no	
560	no	
650	no	
480	no	
104	yes	104
580	no	
204	yes	204
115	yes	115
208	yes	208
152	yes	152
309	no	
432	no	
351	no	
246	no	
182	yes	182
612	no	
256	no	
508	no	
330	no	
274	no	
292	no	
553	no	
529	no	
375	no	

Variant protein HUMCEA_PEA_1_P4 is encoded by the following transcript(s): HUMCEA_PEA_1_T8, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCEA_PEA_1_T8 is shown in bold; this coding portion starts at position 115 and ends at position 1059. The transcript also has the following SNPs as listed in Table 944 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 944 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
49	T ->	No
273	A -> C	Yes
303	T -> G	No
324	T -> C	Yes
352	A -> G	Yes
362	T -> C	Yes
524	A -> C	Yes
631	G -> A	No
1315	A -> G	No
1380	T -> C	No
1533	C -> A	Yes
1706	G -> A	Yes
2308	T -> C	No
2362	C -> T	No
2455	A ->	No
2504	C -> A	Yes
2558	G ->	No
2623	G ->	No

2639	T -> A	No
2640	T -> A	No
2832	G -> A	Yes
2885	C -> T	No
3396	A -> G	Yes
3562	C -> T	Yes
3753	C -> T	Yes

Variant protein HUMCEA_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCEA_PEA_1_T9. An alignment is given to the known protein (Carcinoembryonic antigen-related cell adhesion molecule 5 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCEA_PEA_1_P5 and CEA5_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWV NNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDA PTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTC QAHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWV NNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELSVDHSDPVILNVLYGPDD PTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQELFISNITEKNSGLYTCQ ANNSASGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVN GQSLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTP IISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFV SNLATGRNNSIVKSITVS corresponding to amino acids 1 - 675 of CEA5_HUMAN, which

also corresponds to amino acids 1 - 675 of HUMCEA_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKWLPGASASYSGVESIWFSPKSQEDIFFPSLCSMGTRKSQILS corresponding to amino acids 676 - 719 of HUMCEA_PEA_1_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMCEA_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKWLPGASASYSGVESIWFSPKSQEDIFFPSLCSMGTRKSQILS in HUMCEA_PEA_1_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCEA_PEA_1_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 945, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 945 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
63	F -> L	No
80	I -> V	Yes
83	V -> A	Yes

137	Q -> P	Yes
173	D -> N	No
289	I -> T	No
340	A -> D	Yes
398	E -> K	Yes
647	P ->	No
664	R -> S	Yes

The glycosylation sites of variant protein HUMCEA_PEA_1_P5, as compared to the known protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor, are described in Table 946 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 946 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
197	yes	197
466	yes	466
360	yes	360
288	yes	288
665	yes	665
560	yes	560
650	yes	650
480	yes	480
104	yes	104
580	yes	580
204	yes	204
115	yes	115
208	yes	208

152	yes	152
309	yes	309
432	yes	432
351	yes	351
246	yes	246
182	yes	182
612	yes	612
256	yes	256
508	yes	508
330	yes	330
274	yes	274
292	yes	292
553	yes	553
529	yes	529
375	yes	375

Variant protein HUMCEA_PEA_1_P5 is encoded by the following transcript(s): HUMCEA_PEA_1_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCEA_PEA_1_T9 is shown in bold; this coding portion starts at position 115 and ends at position 2271. The transcript also has the following SNPs as listed in Table 947 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 947 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
49	T ->	No
273	A -> C	Yes
303	T -> G	No

324	T -> C	Yes
352	A -> G	Yes
362	T -> C	Yes
524	A -> C	Yes
631	G -> A	No
915	A -> G	No
980	T -> C	No
1133	C -> A	Yes
1306	G -> A	Yes
1908	T -> C	No
1962	C -> T	No
2055	A ->	No
2104	C -> A	Yes
3259	T -> C	Yes

Variant protein HUMCEA_PEA_1_P14 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCEA_PEA_1_T20. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMCEA_PEA_1_P14 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 948, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 948 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
63	F -> L	No
80	I -> V	Yes
83	V -> A	Yes
137	Q -> P	Yes
173	D -> N	No
289	I -> T	No
340	A -> D	Yes
398	E -> K	Yes

Variant protein HUMCEA_PEA_1_P14 is encoded by the following transcript(s): HUMCEA_PEA_1_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCEA_PEA_1_T20 is shown in bold; this coding portion starts at position 115 and ends at position 1821. The transcript also has the following SNPs as listed in Table 949 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P14 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 949 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
49	T ->	No
273	A -> C	Yes
303	T -> G	No
324	T -> C	Yes
352	A -> G	Yes
362	T -> C	Yes

524	A -> C	Yes
631	G -> A	No
915	A -> G	No
980	T -> C	No
1133	C -> A	Yes
1306	G -> A	Yes

Variant protein HUMCEA_PEA_1_P19 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCEA_PEA_1_T25. An alignment is given to the known protein (Carcinoembryonic antigen-related cell adhesion molecule 5 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCEA_PEA_1_P19 and CEA5_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P19, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYPELPKPSISSNNSKPVEDKDAVAFTCEPETQDATYLWWV NNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVILN corresponding to amino acids 1 - 232 of CEA5_HUMAN, which also corresponds to amino acids 1 - 232 of HUMCEA_PEA_1_P19, and a second amino acid sequence being at least 90 % homologous to VLYGPDTPIISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVALI corresponding to amino acids 589 - 702 of CEA5_HUMAN, which also corresponds to amino acids 233 - 346 of HUMCEA_PEA_1_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HUMCEA_PEA_1_P19, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise NV, having a structure as follows: a sequence starting from any of amino acid numbers 232-x to 232; and ending at any of amino acid numbers 233+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because of manual inspection of known protein localization and/or gene structure.

Variant protein HUMCEA_PEA_1_P19 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 950, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 950 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
63	F -> L	No
80	I -> V	Yes
83	V -> A	Yes
137	Q -> P	Yes
173	D -> N	No
291	P ->	No
308	R -> S	Yes

326	G ->	No

The glycosylation sites of variant protein HUMCEA_PEA_1_P19, as compared to the known protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor, are described in Table 951 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 951 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
197	yes	197
466	no	
360	no	
288	no	
665	yes	309
560	no	
650	yes	294
480	no	
104	yes	104
580	no	
204	yes	204
115	yes	115
208	yes	208
152	yes	152
309	no	
432	no	
351	no	
246	no	
182	yes	182

612	yes	256
256	no	
508	no	
330	no	
274	no	
292	no	
553	no	
529	no	
375	no	

Variant protein HUMCEA_PEA_1_P19 is encoded by the following transcript(s): HUMCEA_PEA_1_T25, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCEA_PEA_1_T25 is shown in bold; this coding portion starts at position 115 and ends at position 1152. The transcript also has the following SNPs as listed in Table 952 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 952 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
49	T ->	No
273	A -> C	Yes
303	T -> G	No
324	T -> C	Yes
352	A -> G	Yes
362	T -> C	Yes
524	A -> C	Yes
631	G -> A	No
840	T -> C	No

894	C -> T	No
987	A ->	No
1036	C -> A	Yes
1090	G ->	No
1155	G ->	No
1171	T -> A	No
1172	T -> A	No
1364	G -> A	Yes
1417	C -> T	No
1928	A -> G	Yes
2094	C -> T	Yes
2285	C -> T	Yes

Variant protein HUMCEA_PEA_1_P20 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMCEA_PEA_1_T26. An alignment is given to the known protein (Carcinoembryonic antigen-related cell adhesion molecule 5 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMCEA_PEA_1_P20 and CEA5_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMCEA_PEA_1_P20, comprising a first amino acid sequence being at least 90 % homologous to MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQ HLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYT LHVIKSDLVNEEATGQFRVYP corresponding to amino acids 1 - 142 of CEA5_HUMAN, which also corresponds to amino acids 1 - 142 of HUMCEA_PEA_1_P20, and a second amino acid sequence being at least 90 % homologous to ELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLT LFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGANLNLSCHS

ASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASG TSPGLSAGATVGIMIGVLVGVALI corresponding to amino acids 499 - 702 of CEA5_HUMAN, which also corresponds to amino acids 143 - 346 of HUMCEA_PEA_1_P20, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HUMCEA_PEA_1_P20, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise PE, having a structure as follows: a sequence starting from any of amino acid numbers 142-x to 142; and ending at any of amino acid numbers 143+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because of manual inspection of known protein localization and/or gene structure.

Variant protein HUMCEA_PEA_1_P20 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 953, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P20 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 953 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
63	F -> L	No
80	I -> V	Yes

83	V -> A	Yes
137	Q -> P	Yes
291	P ->	No
308	R -> S	Yes
326	G ->	No

The glycosylation sites of variant protein HUMCEA_PEA_1_P20, as compared to the known protein Carcinoembryonic antigen-related cell adhesion molecule 5 precursor, are described in Table 954 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 954 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
197	no	
466	no	
360	no	
288	no	
665	yes	309
560	yes	204
650	yes	294
480	no	
104	yes	104
580	yes	224
204	no	
115	yes	115
208	no	
152	no	
309	no	

432	no	
351	no	
246	no	
182	no	
612	yes	256
256	no	
508	yes	152
330	no	
274	no	
292	no	
553	yes	197
529	yes	173
375	no	

Variant protein HUMCEA_PEA_1_P20 is encoded by the following transcript(s): HUMCEA_PEA_1_T26, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMCEA_PEA_1_T26 is shown in bold; this coding portion starts at position 115 and ends at position 1152. The transcript also has the following SNPs as listed in Table 955 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMCEA_PEA_1_P20 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 955 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
49	T ->	No
273	A -> C	Yes
303	T -> G	No
324	T -> C	Yes
352	A -> G	Yes

362	T -> C	Yes
524	A -> C	Yes
840	T -> C	No
894	C -> T	No
987	A ->	No
1036	C -> A	Yes
1090	G ->	No
1155	G ->	No
1171	T -> A	No
1172	T -> A	No
1364	G -> A	Yes
1417	C -> T	No
1928	A -> G	Yes
2094	C -> T	Yes
2285	C -> T	Yes

As noted above, cluster HUMCEA features 42 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMCEA_PEA_1_node_0 according to the present invention is supported by 56 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 956 below describes the starting and ending position of this segment on each transcript.

Table 956 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HUMCEA_PEA_1_T8	1	178
HUMCEA_PEA_1_T9	1	178
HUMCEA_PEA_1_T20	1	178
HUMCEA_PEA_1_T25	1	178
HUMCEA_PEA_1_T26	1	178

Segment cluster HUMCEA_PEA_1_node_2 according to the present invention is supported by 83 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 957 below describes the starting and ending position of this segment on each transcript.

Table 957 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	179	456
HUMCEA_PEA_1_T9	179	456
HUMCEA_PEA_1_T20	179	456
HUMCEA_PEA_1_T25	179	456
HUMCEA_PEA_1_T26	179	456

Segment cluster HUMCEA_PEA_1_node_11 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8. Table 958 below describes the starting and ending position of this segment on each transcript.

Table 958 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	818	1217

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 959.

Table 959 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
HUMCEA_0_0_96	lung malignant tumors	LUN

Segment cluster HUMCEA_PEA_1_node_12 according to the present invention is supported by 83 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 960 below describes the starting and ending position of this segment on each transcript.

Table 960 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1218	1472
HUMCEA_PEA_1_T9	818	1072
HUMCEA_PEA_1_T20	818	1072

Segment cluster HUMCEA_PEA_1_node_31 according to the present invention is supported by 87 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8,

HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 961 below describes the starting and ending position of this segment on each transcript.

Table 961 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1817	2006
HUMCEA_PEA_1_T9	1417	1606
HUMCEA_PEA_1_T20	1417	1606

Segment cluster HUMCEA_PEA_1_node_36 according to the present invention is supported by 94 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T26. Table 962 below describes the starting and ending position of this segment on each transcript.

Table 962 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2159	2285
HUMCEA_PEA_1_T9	1759	1885
HUMCEA_PEA_1_T26	691	817

Segment cluster HUMCEA_PEA_1_node_44 according to the present invention is supported by 112 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 963 below describes the starting and ending position of this segment on each transcript.

Table 963 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2286	2540
HUMCEA_PEA_1_T9	1886	2140
HUMCEA_PEA_1_T25	818	1072
HUMCEA_PEA_1_T26	818	1072

Segment cluster HUMCEA_PEA_1_node_46 according to the present invention is supported by 15 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T9. Table 964 below describes the starting and ending position of this segment on each transcript.

Table 964 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T9	2174	3347

Segment cluster HUMCEA_PEA_1_node_63 according to the present invention is supported by 68 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 965 below describes the starting and ending position of this segment on each transcript.

Table 965 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2957	3135
HUMCEA_PEA_1_T25	1489	1667
HUMCEA_PEA_1_T26	1489	1667

Segment cluster HUMCEA_PEA_1_node_65 according to the present invention is supported by 54 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 966 below describes the starting and ending position of this segment on each transcript.

Table 966 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	3166	3897
HUMCEA_PEA_1_T25	1698	2429
HUMCEA_PEA_1_T26	1698	2429

Segment cluster HUMCEA_PEA_1_node_67 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T20. Table 967 below describes the starting and ending position of this segment on each transcript.

Table 967 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T20	1607	1886

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMCEA_PEA_1_node_3 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 968 below describes the starting and ending position of this segment on each transcript.

Table 968 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	457	538
HUMCEA_PEA_1_T9	457	538
HUMCEA_PEA_1_T20	457	538
HUMCEA_PEA_1_T25	457	538
HUMCEA_PEA_1_T26	457	538

Segment cluster HUMCEA_PEA_1_node_7 according to the present invention is supported by 73 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20 and HUMCEA_PEA_1_T25. Table 969 below describes the starting and ending position of this segment on each transcript.

Table 969 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	539	642
HUMCEA_PEA_1_T9	539	642
HUMCEA_PEA_1_T20	539	642
HUMCEA_PEA_1_T25	539	642

Segment cluster HUMCEA_PEA_1_node_8 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20 and HUMCEA_PEA_1_T25. Table 970 below describes the starting and ending position of this segment on each transcript.

Table 970- - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	643	690
HUMCEA_PEA_1_T9	643	690
HUMCEA_PEA_1_T20	643	690
HUMCEA_PEA_1_T25	643	690

Segment cluster HUMCEA_PEA_1_node_9 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20 and HUMCEA_PEA_1_T25. Table 971 below describes the starting and ending position of this segment on each transcript.

Table 971 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	691	738
HUMCEA_PEA_1_T9	691	738
HUMCEA_PEA_1_T20	691	738
HUMCEA_PEA_1_T25	691	738

Segment cluster HUMCEA_PEA_1_node_10 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This

segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9, HUMCEA_PEA_1_T20 and HUMCEA_PEA_1_T25. Table 972 below describes the starting and ending position of this segment on each transcript.

Table 972 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	739	817
HUMCEA_PEA_1_T9	739	817
HUMCEA_PEA_1_T20	739	817
HUMCEA_PEA_1_T25	739	817

Segment cluster HUMCEA_PEA_1_node_15 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 973 below describes the starting and ending position of this segment on each transcript.

Table 973 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1473	1475
HUMCEA_PEA_1_T9	1073	1075
HUMCEA_PEA_1_T20	1073	1075

Segment cluster HUMCEA_PEA_1_node_16 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 974 below describes the starting and ending position of this segment on each transcript.

Table 974 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1476	1481
HUMCEA_PEA_1_T9	1076	1081
HUMCEA_PEA_1_T20	1076	1081

Segment cluster HUMCEA_PEA_1_node_17 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 975 below describes the starting and ending position of this segment on each transcript.

Table 975 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1482	1488
HUMCEA_PEA_1_T9	1082	1088
HUMCEA_PEA_1_T20	1082	1088

Segment cluster HUMCEA_PEA_1_node_18 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 976 below describes the starting and ending position of this segment on each transcript.

Table 976 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1489	1506
HUMCEA_PEA_1_T9	1089	1106
HUMCEA_PEA_1_T20	1089	1106

Segment cluster HUMCEA_PEA_1_node_19 according to the present invention is supported by 69 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 977 below describes the starting and ending position of this segment on each transcript.

Table 977 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1507	1576
HUMCEA_PEA_1_T9	1107	1176
HUMCEA_PEA_1_T20	1107	1176

Segment cluster HUMCEA_PEA_1_node_20 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 978 below describes the starting and ending position of this segment on each transcript.

Table 978 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1577	1600
HUMCEA_PEA_1_T9	1177	1200
HUMCEA_PEA_1_T20	1177	1200

Segment cluster HUMCEA_PEA_1_node_21 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and

HUMCEA_PEA_1_T20. Table 979 below describes the starting and ending position of this segment on each transcript.

Table 979 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1601	1624
HUMCEA_PEA_1_T9	1201	1224
HUMCEA_PEA_1_T20	1201	1224

Segment cluster HUMCEA_PEA_1_node_22 according to the present invention is supported by 77 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 980 below describes the starting and ending position of this segment on each transcript.

Table 980 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1625	1702
HUMCEA_PEA_1_T9	1225	1302
HUMCEA_PEA_1_T20	1225	1302

Segment cluster HUMCEA_PEA_1_node_23 according to the present invention is supported by 72 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 981 below describes the starting and ending position of this segment on each transcript.

Table 981 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1703	1732
HUMCEA_PEA_1_T9	1303	1332
HUMCEA_PEA_1_T20	1303	1332

Segment cluster HUMCEA_PEA_1_node_24 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 982 below describes the starting and ending position of this segment on each transcript.

Table 982 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1733	1751
HUMCEA_PEA_1_T9	1333	1351
HUMCEA_PEA_1_T20	1333	1351

Segment cluster HUMCEA_PEA_1_node_27 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 983 below describes the starting and ending position of this segment on each transcript.

Table 983 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1752	1770
HUMCEA_PEA_1_T9	1352	1370
HUMCEA_PEA_1_T20	1352	1370

Segment cluster HUMCEA_PEA_1_node_29 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 984 below describes the starting and ending position of this segment on each transcript.

Table 984 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1771	1788
HUMCEA_PEA_1_T9	1371	1388
HUMCEA_PEA_1_T20	1371	1388

Segment cluster HUMCEA_PEA_1_node_30 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T20. Table 985 below describes the starting and ending position of this segment on each transcript.

Table 985 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	1789	1816
HUMCEA_PEA_1_T9	1389	1416
HUMCEA_PEA_1_T20	1389	1416

Segment cluster HUMCEA_PEA_1_node_33 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and

HUMCEA_PEA_1_T26. Table 986 below describes the starting and ending position of this segment on each transcript.

Table 986 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2007	2028
HUMCEA_PEA_1_T9	1607	1628
HUMCEA_PEA_1_T26	539	560

Segment cluster HUMCEA_PEA_1_node_34 according to the present invention is supported by 80 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T26. Table 987 below describes the starting and ending position of this segment on each transcript.

Table 987 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2029	2110
HUMCEA_PEA_1_T9	1629	1710
HUMCEA_PEA_1_T26	561	642

Segment cluster HUMCEA_PEA_1_node_35 according to the present invention is supported by 75 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T9 and HUMCEA_PEA_1_T26. Table 988 below describes the starting and ending position of this segment on each transcript.

Table 988 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2111	2158
HUMCEA_PEA_1_T9	1711	1758
HUMCEA_PEA_1_T26	643	690

Segment cluster HUMCEA_PEA_1_node_45 according to the present invention is supported by 9 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T9. Table 989 below describes the starting and ending position of this segment on each transcript.

Table 989 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T9	2141	2173

Segment cluster HUMCEA_PEA_1_node_50 according to the present invention is supported by 64 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 990 below describes the starting and ending position of this segment on each transcript.

Table 990 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2541	2567
HUMCEA_PEA_1_T25	1073	1099
HUMCEA_PEA_1_T26	1073	1099

Segment cluster HUMCEA_PEA_1_node_51 according to the present invention is supported by 88 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 991 below describes the starting and ending position of this segment on each transcript.

Table 991 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2568	2659
HUMCEA_PEA_1_T25	1100	1191
HUMCEA_PEA_1_T26	1100	1191

Segment cluster HUMCEA_PEA_1_node_56 according to the present invention is supported by 75 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 992 below describes the starting and ending position of this segment on each transcript.

Table 992 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2660	2685
HUMCEA_PEA_1_T25	1192	1217
HUMCEA_PEA_1_T26	1192	1217

Segment cluster HUMCEA_PEA_1_node_57 according to the present invention is supported by 82 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8,

HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 993 below describes the starting and ending position of this segment on each transcript.

Table 993 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2686	2786
HUMCEA_PEA_1_T25	1218	1318
HUMCEA_PEA_1_T26	1218	1318

Segment cluster HUMCEA_PEA_1_node_58 according to the present invention is supported by 63 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 994 below describes the starting and ending position of this segment on each transcript.

Table 994 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2787	2820
HUMCEA_PEA_1_T25	1319	1352
HUMCEA_PEA_1_T26	1319	1352

Segment cluster HUMCEA_PEA_1_node_60 according to the present invention is supported by 55 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 995 below describes the starting and ending position of this segment on each transcript.

Table 995 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2821	2864
HUMCEA_PEA_1_T25	1353	1396
HUMCEA_PEA_1_T26	1353	1396

Segment cluster HUMCEA_PEA_1_node_61 according to the present invention can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 996 below describes the starting and ending position of this segment on each transcript.

Table 996 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2865	2868
HUMCEA_PEA_1_T25	1397	1400
HUMCEA_PEA_1_T26	1397	1400

Segment cluster HUMCEA_PEA_1_node_62 according to the present invention is supported by 60 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 997 below describes the starting and ending position of this segment on each transcript.

Table 997 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	2869	2956
HUMCEA_PEA_1_T25	1401	1488

HUMCEA_PEA_1_T26	1401	1488

Segment cluster HUMCEA_PEA_1_node_64 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMCEA_PEA_1_T8, HUMCEA_PEA_1_T25 and HUMCEA_PEA_1_T26. Table 998 below describes the starting and ending position of this segment on each transcript.

Table 998 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMCEA_PEA_1_T8	3136	3165
HUMCEA_PEA_1_T25	1668	1697
HUMCEA_PEA_1_T26	1668	1697

Variant protein alignment to the previously known protein:

Sequence name: CEA5_HUMAN

Sequence documentation:

Alignment of: HUMCEA_PEA_1_P4 x CEA5_HUMAN ..

Alignment segment 1/1:

Quality: 2320.00

Escore: 0

1016

Matching length: 234 Total length: 234

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
          .         .         .         .         .
1   MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1   MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
          .         .         .         .         .
51  VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
51  VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
          .         .         .         .         .
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
          .         .         .         .         .
151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
          .         .         .
201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVL                  234
    |||||||||||||||||||||||||||||||||
201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVL                  234
```

Sequence name: CEA5_HUMAN

Sequence documentation:

Alignment of: HUMCEA_PEA_1_P5 x CEA5_HUMAN ..

Alignment segment 1/1:

Quality: 6692.00

Escore: 0

Matching length: 675 Total length: 675

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
                .         .         .         .         .
       51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
```

```
                .         .         .         .         .
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
                .         .         .         .         .
151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
                .         .         .         .         .
201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPLNTSYR 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPLNTSYR 250
                .         .         .         .         .
251 SGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQ 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 SGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQ 300
                .         .         .         .         .
301 AHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQ 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 AHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQ 350
                .         .         .         .         .
351 NTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELS 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 NTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELS 400
                .         .         .         .         .
401 VDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWL 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 VDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWL 450
                .         .         .         .         .
451 IDGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAEL 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
451 IDGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAEL 500
                .         .         .         .         .
501 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLS 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLS 550
                .         .         .         .         .
551 NGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
551 NGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP 600
                .         .         .         .         .
601 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
601 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN 650
                .         .
651 GTYACFVSNLATGRNNSIVKSITVS                          675
    |||||||||||||||||||||||||
651 GTYACFVSNLATGRNNSIVKSITVS                          675
```

Sequence name: CEA5_HUMAN

Sequence documentation:

Alignment of: HUMCEA_PEA_1_P19 x CEA5_HUMAN ..

Alignment segment 1/1:

1020

Quality: 3298.00

Escore: 0

Matching length: 346 Total length: 702

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 49.29 Total Percent Identity: 49.29

Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
                .         .         .         .         .
     51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
                .         .         .         .         .
    101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
                .         .         .         .         .
    151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
                .         .         .         .         .
    201 TLFNVTRNDTASYKCETQNPVSARRSDSVILN..................  232
        ||||||||||||||||||||||||||||||
    201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPLNTSYR 250
```

```
                .         .         .         .         .
          232 ..................................................  232

          251 SGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQ  300
                .         .         .         .         .
          232 ..................................................  232

          301 AHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQ  350
                .         .         .         .         .
          232 ..................................................  232

          351 NTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELS  400
                .         .         .         .         .
          232 ..................................................  232

          401 VDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWL  450
                .         .         .         .         .
          232 ..................................................  232

          451 IDGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAEL  500
                .         .         .         .         .
          232 ..................................................  232

          501 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLS  550
                .         .         .         .         .
          233 ....................................VLYGPDTPIISP  244
                                                  ||||||||||||
          551 NGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP  600
                .         .         .         .         .
          245 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN  294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
601 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN 650
                .         .         .         .         .
295 GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVA 344
    |||||||||||||||||||||||||||||||||||||||||||||||||||
651 GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVA 700

345 LI                                                  346
    ||
701 LI                                                  702
```

Sequence name: CEA5_HUMAN

Sequence documentation:

Alignment of: HUMCEA_PEA_1_P20 x CEA5_HUMAN ..

Alignment segment 1/1:

Quality: 3294.00

Escore: 0

Matching length: 346 Total length: 702

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 49.29 Total Percent Identity: 49.29

1023

Gaps: 1

Alignment:

```
                 .         .         .         .         .
1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
  ||||||||||||||||||||||||||||||||||||||||||||||||||
1 MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKE 50
                 .         .         .         .         .
51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
  ||||||||||||||||||||||||||||||||||||||||||||||||||
51 VLLLVHNLPQHLFGYSWYKGERVDGNRQIIGYVIGTQQATPGPAYSGREI 100
                 .         .         .         .         .
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYP..... 142
  |||||||||||||||||||||||||||||||||||||||||
101 IYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSIS 150
                 .         .         .         .         .
142 ..................................................... 142

151 SNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTL 200
                 .         .         .         .         .
142 ..................................................... 142

201 TLFNVTRNDTASYKCETQNPVSARRSDSVILNVLYGPDAPTISPLNTSYR 250
                 .         .         .         .         .
142 ..................................................... 142

251 SGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQ 300
                 .         .         .         .         .
142 ..................................................... 142

301 AHNSDTGLNRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQ 350
```

```
                .         .         .         .         .
142 ..................................................  142

351 NTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPYECGIQNELS  400
                .         .         .         .         .
142 ..................................................  142

401 VDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWL  450
                .         .         .         .         .
143 ..............................................EL  144
                                                   ||
451 IDGNIQQHTQELFISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAEL  500
                .         .         .         .         .
145 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLS  194
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 PKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLS  550
                .         .         .         .         .
195 NGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP  244
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 NGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISP  600
                .         .         .         .         .
245 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN  294
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 PDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNN  650
                .         .         .         .         .
295 GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVA  344
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 GTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVA  700

345 LI                                                  346
    ||
```

1025

701 LI 702

DESCRIPTION FOR CLUSTER R35137

Cluster R35137 features 6 transcript(s) and 20 segment(s) of interest, the names for which are given in Tables 999 and 1000, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1001.

Table 999 - Transcripts of interest

Transcript Name	Sequence ID No.
R35137_PEA_1_PEA_1_PEA_1_T3	114
R35137_PEA_1_PEA_1_PEA_1_T5	115
R35137_PEA_1_PEA_1_PEA_1_T10	116
R35137_PEA_1_PEA_1_PEA_1_T11	117
R35137_PEA_1_PEA_1_PEA_1_T12	118
R35137_PEA_1_PEA_1_PEA_1_T14	119

Table 1000 - Segments of interest

Segment Name	Sequence ID No.
R35137_PEA_1_PEA_1_PEA_1_node_2	856
R35137_PEA_1_PEA_1_PEA_1_node_3	857
R35137_PEA_1_PEA_1_PEA_1_node_9	858
R35137_PEA_1_PEA_1_PEA_1_node_11	859
R35137_PEA_1_PEA_1_PEA_1_node_16	860
R35137_PEA_1_PEA_1_PEA_1_node_18	861
R35137_PEA_1_PEA_1_PEA_1_node_20	862
R35137_PEA_1_PEA_1_PEA_1_node_27	863
R35137_PEA_1_PEA_1_PEA_1_node_5	864
R35137_PEA_1_PEA_1_PEA_1_node_7	865
R35137_PEA_1_PEA_1_PEA_1_node_12	866

R35137_PEA_1_PEA_1_PEA_1_node_14	867
R35137_PEA_1_PEA_1_PEA_1_node_15	868
R35137_PEA_1_PEA_1_PEA_1_node_17	869
R35137_PEA_1_PEA_1_PEA_1_node_21	870
R35137_PEA_1_PEA_1_PEA_1_node_22	871
R35137_PEA_1_PEA_1_PEA_1_node_23	872
R35137_PEA_1_PEA_1_PEA_1_node_24	873
R35137_PEA_1_PEA_1_PEA_1_node_25	874
R35137_PEA_1_PEA_1_PEA_1_node_26	875

Table 1001 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
R35137_PEA_1_PEA_1_PEA_1_P9	1385	R35137_PEA_1_PEA_1_PEA_1_T10; R35137_PEA_1_PEA_1_PEA_1_T12
R35137_PEA_1_PEA_1_PEA_1_P8	1386	R35137_PEA_1_PEA_1_PEA_1_T11
R35137_PEA_1_PEA_1_PEA_1_P11	1387	R35137_PEA_1_PEA_1_PEA_1_T14
R35137_PEA_1_PEA_1_PEA_1_P2	1388	R35137_PEA_1_PEA_1_PEA_1_T3
R35137_PEA_1_PEA_1_PEA_1_P4	1389	R35137_PEA_1_PEA_1_PEA_1_T5

These sequences are variants of the known protein Alanine aminotransferase (SwissProt accession identifier ALAT_HUMAN; known also according to the synonyms EC 2.6.1.2; Glutamic--pyruvic transaminase; GPT; Glutamic--alanine transaminase), SEQ ID NO: 1452, referred to herein as the previously known protein.

Protein Alanine aminotransferase is known or believed to have the following function(s): Participates in cellular nitrogen metabolism and also in liver gluconeogenesis starting with precursors transported from skeletal muscles. The sequence for protein Alanine aminotransferase is given at the end of the application, as "Alanine aminotransferase amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1002.

Table 1002 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
13	H -> N (in allele GPT*2; dbSNP:1063739). /FTId=VAR_000561.
3 - 6	STGD -> RRGN
38	G -> S
221	A -> H

Protein Alanine aminotransferase localization is believed to be Cytoplasmic.

Cluster R35137 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 34 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 34 and Table 1003. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: hepatocellular carcinoma.

Table 1003 - Normal tissue distribution

Name of Tissue	Number
brain	12
epithelial	16
general	8
kidney	20
liver	0
lung	0
pancreas	2

prostate	0

Table 1004 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
brain	3.2e-01	4.8e-01	1.8e-01	2.5	4.2e-01	1.5
epithelial	7.6e-01	7.7e-01	8.9e-01	0.5	9.8e-01	0.4
general	6.7e-01	8.2e-01	4.2e-01	1.0	8.5e-01	0.7
kidney	8.6e-01	9.0e-01	5.8e-01	0.9	7.0e-01	0.8
liver	1.8e-01	4.5e-01	3.0e-03	7.6	1.6e-01	2.3
lung	1	6.3e-01	1	1.0	6.2e-01	1.6
pancreas	2.3e-01	4.0e-01	1.8e-01	3.1	2.8e-01	2.3
prostate	1	7.8e-01	1	1.0	7.5e-01	1.3

As noted above, cluster R35137 features 6 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Alanine aminotransferase. A description of each variant protein according to the present invention is now provided.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R35137_PEA_1_PEA_1_PEA_1_T10. An alignment is given to the known protein (Alanine aminotransferase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R35137_PEA_1_PEA_1_PEA_1_P9 and ALAT_HUMAN_V1 (SEQ ID NO: 1453):

1.An isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P9, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG

GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEV corresponding to amino acids 1 - 274 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 274 of R35137_PEA_1_PEA_1_PEA_1_P9, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRAYEAGGGSRAMARPSSPDGPPPPPHLTWPCAGAGSAAAMWRW corresponding to amino acids 275 - 385 of R35137_PEA_1_PEA_1_PEA_1_P9, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRAYEAGGGSRAMARPSSPDGPPPPPHLTWPCAGAGSAAAMWRW in R35137_PEA_1_PEA_1_PEA_1_P9.

It should be noted that the known protein sequence (ALAT_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ALAT_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 1005 - Changes to ALAT_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met
222	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P9 is encoded by the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R35137_PEA_1_PEA_1_PEA_1_T10 is shown in bold; this coding portion starts at position 271 and ends at position 1425. The transcript also has the following SNPs as listed in Table 1006 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1006 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
230	C -> T	No
231	C -> T	No
310	C -> A	Yes
432	G ->	No
969	C ->	No
1225	G ->	No
1745	T -> G	No
1957	C ->	No
2018	G -> A	No

2019	C -> A	No
2101	A -> G	No
2102	A -> G	No
2159	C -> T	Yes
2710	G -> C	No
2789	C -> A	Yes
3622	G -> A	Yes

Variant protein R35137_PEA_1_PEA_1_PEA_1_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R35137_PEA_1_PEA_1_PEA_1_T11. An alignment is given to the known protein (Alanine aminotransferase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R35137_PEA_1_PEA_1_PEA_1_P8 and ALAT_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P8, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVL MEMGPPYAGQQELASFHSTSKGYMGEC corresponding to amino acids 1 - 320 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 320 of R35137_PEA_1_PEA_1_PEA_1_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRTRRVGARGPWPGPPRPMGHPLLRT corresponding to amino acids 321 - 346 of

R35137_PEA_1_PEA_1_PEA_1_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRTRRVGARGPWPGPPRPMGHPLLRT in R35137_PEA_1_PEA_1_PEA_1_P8.

It should be noted that the known protein sequence (ALAT_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ALAT_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 1007 - Changes to ALAT_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met
222	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1008, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein

R35137_PEA_1_PEA_1_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1008 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	H -> N	Yes
54	Q ->	No
233	R ->	No
296	M ->	No

Variant protein R35137_PEA_1_PEA_1_PEA_1_P8 is encoded by the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R35137_PEA_1_PEA_1_PEA_1_T11 is shown in bold; this coding portion starts at position 271 and ends at position 1308. The transcript also has the following SNPs as listed in Table 1009 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1009 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
230	C -> T	No
231	C -> T	No
310	C -> A	Yes
432	G ->	No
969	C ->	No
1158	G ->	No
1752	T -> G	No

2030	C ->	No
2091	G -> A	No
2092	C -> A	No
2174	A -> G	No
2175	A -> G	No
2232	C -> T	Yes
2783	G -> C	No
2862	C -> A	Yes
3695	G -> A	Yes

Variant protein R35137_PEA_1_PEA_1_PEA_1_P11 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R35137_PEA_1_PEA_1_PEA_1_T14. An alignment is given to the known protein (Alanine aminotransferase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R35137_PEA_1_PEA_1_PEA_1_P11 and ALAT_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P11, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQAR corresponding to amino acids 1 - 229 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 229 of R35137_PEA_1_PEA_1_PEA_1_P11, and a second amino acid sequence being at least 90 % homologous to SGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS corresponding to amino acids 455 - 496 of ALAT_HUMAN_V1, which also corresponds to

amino acids 230 - 271 of R35137_PEA_1_PEA_1_PEA_1_P11, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of R35137_PEA_1_PEA_1_PEA_1_P11, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise RS, having a structure as follows: a sequence starting from any of amino acid numbers 229-x to 229; and ending at any of amino acid numbers 230+ ((n-2) - x), in which x varies from 0 to n-2.

It should be noted that the known protein sequence (ALAT_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ALAT_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 1010 - Changes to ALAT_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met
222	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P11 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1011, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1011 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	H -> N	Yes
54	Q ->	No

Variant protein R35137_PEA_1_PEA_1_PEA_1_P11 is encoded by the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T14, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R35137_PEA_1_PEA_1_PEA_1_T14 is shown in bold; this coding portion starts at position 271 and ends at position 1083. The transcript also has the following SNPs as listed in Table 1012 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P11 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1012 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
230	C -> T	No
231	C -> T	No
310	C -> A	Yes
432	G ->	No
1115	C ->	No

1176	G -> A	No
1177	C -> A	No

Variant protein R35137_PEA_1_PEA_1_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R35137_PEA_1_PEA_1_PEA_1_T3. An alignment is given to the known protein (Alanine aminotransferase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R35137_PEA_1_PEA_1_PEA_1_P2 and ALAT_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P2, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEV corresponding to amino acids 1 - 274 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 274 of R35137_PEA_1_PEA_1_PEA_1_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRVPRRLCGGGEHGRCSAAADAEADECAAVPAGARTGPAGPGGQPAR AHRPLLCAVPG corresponding to amino acids 275 - 399 of R35137_PEA_1_PEA_1_PEA_1_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least

about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAVPLIQEGAHGDGAALRRAAGACLLP LHLQGLHGRVRVPRRLCGGGEHGRCSAAADAEADECAAVPAGARTGPAGPGGGQPAR AHRPLLCAVPG in R35137_PEA_1_PEA_1_PEA_1_P2.

It should be noted that the known protein sequence (ALAT_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ALAT_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 1013 - Changes to ALAT_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met
222	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein. In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1014, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1014 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	H -> N	Yes
54	Q ->	No
233	R ->	No
319	G ->	No

Variant protein R35137_PEA_1_PEA_1_PEA_1_P2 is encoded by the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R35137_PEA_1_PEA_1_PEA_1_T3 is shown in bold; this coding portion starts at position 271 and ends at position 1467. The transcript also has the following SNPs as listed in Table 1015 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1015 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
230	C -> T	No
231	C -> T	No
310	C -> A	Yes
432	G ->	No
969	C ->	No
1225	G ->	No
1645	T -> G	No
1857	C ->	No
1918	G -> A	No

1040

1919	C -> A	No
2001	A -> G	No
2002	A -> G	No
2059	C -> T	Yes
2610	G -> C	No
2689	C -> A	Yes
3522	G -> A	Yes

Variant protein R35137_PEA_1_PEA_1_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R35137_PEA_1_PEA_1_PEA_1_T5. An alignment is given to the known protein (Alanine aminotransferase) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R35137_PEA_1_PEA_1_PEA_1_P4 and ALAT_HUMAN_V1:

1.An isolated chimeric polypeptide encoding for R35137_PEA_1_PEA_1_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVK KPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACG GHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVFLSTGASDAIVTVLKLLVAGEG HTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDHCRP RALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVL MEMGPPYAGQQELASFHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRL CPPVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAELAAKAKLTEQVFNEAPGISCNP VQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGICVVPGSGFGQREGTYH FRMTILPPLEKLRLLLEKLSRFHAKFTLE corresponding to amino acids 1 - 494 of ALAT_HUMAN_V1, which also corresponds to amino acids 1 - 494 of R35137_PEA_1_PEA_1_PEA_1_P4, and a second amino acid sequence being at least 70%,

optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPGRLWSPLYLLLMPGGVGWGGCWAPASLQVPNKAVWQSDSKKEALAAAWPAPTCL PFLQA corresponding to amino acids 495 - 555 of R35137_PEA_1_PEA_1_PEA_1_P4, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R35137_PEA_1_PEA_1_PEA_1_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPGRLWSPLYLLLMPGGVGWGGCWAPASLQVPNKAVWQSDSKKEALAAAWPAPTCL PFLQA in R35137_PEA_1_PEA_1_PEA_1_P4.

It should be noted that the known protein sequence (ALAT_HUMAN) has one or more changes than the sequence given at the end of the application and named as being the amino acid sequence for ALAT_HUMAN_V1. These changes were previously known to occur and are listed in the table below.

Table 1016 - Changes to ALAT_HUMAN_V1

SNP position(s) on amino acid sequence	Type of change
1	init_met
222	conflict

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: intracellularly. The protein localization is believed to be intracellularly because neither of the trans-membrane region prediction programs predicted a trans-membrane region for this protein.

In addition both signal-peptide prediction programs predict that this protein is a non-secreted protein.

Variant protein R35137_PEA_1_PEA_1_PEA_1_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1017, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1017 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
14	H -> N	Yes
54	Q ->	No
233	R ->	No
296	M ->	No
436	D -> E	No
508	M -> I	No
509	P -> T	No
536	K -> R	No

Variant protein R35137_PEA_1_PEA_1_PEA_1_P4 is encoded by the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R35137_PEA_1_PEA_1_PEA_1_T5 is shown in bold; this coding portion starts at position 271 and ends at position 1935. The transcript also has the following SNPs as listed in Table 1018 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R35137_PEA_1_PEA_1_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1018 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
230	C -> T	No
231	C -> T	No
310	C -> A	Yes
432	G ->	No
969	C ->	No
1158	G ->	No
1578	T -> G	No
1794	G -> A	No
1795	C -> A	No
1877	A -> G	No
1878	A -> G	No
1935	C -> T	Yes
2486	G -> C	No
2565	C -> A	Yes
3398	G -> A	Yes

As noted above, cluster R35137 features 20 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_2 according to the present invention is supported by 19 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1019 below describes the starting and ending position of this segment on each transcript.

Table 1019 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1	266
R35137_PEA_1_PEA_1_PEA_1_T5	1	266
R35137_PEA_1_PEA_1_PEA_1_T10	1	266
R35137_PEA_1_PEA_1_PEA_1_T11	1	266
R35137_PEA_1_PEA_1_PEA_1_T12	1	266
R35137_PEA_1_PEA_1_PEA_1_T14	1	266

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_3 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1020 below describes the starting and ending position of this segment on each transcript.

Table 1020 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	267	432
R35137_PEA_1_PEA_1_PEA_1_T5	267	432
R35137_PEA_1_PEA_1_PEA_1_T10	267	432
R35137_PEA_1_PEA_1_PEA_1_T11	267	432
R35137_PEA_1_PEA_1_PEA_1_T12	267	432
R35137_PEA_1_PEA_1_PEA_1_T14	267	432

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_9 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1021 below describes the starting and ending position of this segment on each transcript.

Table 1021 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	632	765
R35137_PEA_1_PEA_1_PEA_1_T5	632	765
R35137_PEA_1_PEA_1_PEA_1_T10	632	765
R35137_PEA_1_PEA_1_PEA_1_T11	632	765
R35137_PEA_1_PEA_1_PEA_1_T12	632	765
R35137_PEA_1_PEA_1_PEA_1_T14	632	765

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_11 according to the present invention is supported by 30 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1022 below describes the starting and ending position of this segment on each transcript.

Table 1022 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	766	955

R35137_PEA_1_PEA_1_PEA_1_T5	766	955
R35137_PEA_1_PEA_1_PEA_1_T10	766	955
R35137_PEA_1_PEA_1_PEA_1_T11	766	955
R35137_PEA_1_PEA_1_PEA_1_T12	766	955
R35137_PEA_1_PEA_1_PEA_1_T14	766	955

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_16 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1023 below describes the starting and ending position of this segment on each transcript.

Table 1023 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1157	1293
R35137_PEA_1_PEA_1_PEA_1_T5	1090	1226
R35137_PEA_1_PEA_1_PEA_1_T10	1157	1293
R35137_PEA_1_PEA_1_PEA_1_T11	1090	1226
R35137_PEA_1_PEA_1_PEA_1_T12	1157	1293

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_18 according to the present invention is supported by 24 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and

R35137_PEA_1_PEA_1_PEA_1_T12. Table 1024 below describes the starting and ending position of this segment on each transcript.

Table 1024 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1294	1468
R35137_PEA_1_PEA_1_PEA_1_T5	1227	1401
R35137_PEA_1_PEA_1_PEA_1_T10	1394	1568
R35137_PEA_1_PEA_1_PEA_1_T11	1327	1501
R35137_PEA_1_PEA_1_PEA_1_T12	1394	1568

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment (in relation to lung cancer), shown in Table 1025.

Table 1025 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
R35137_0_5_0	lung malignant tumors	LUN

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_20 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1026 below describes the starting and ending position of this segment on each transcript.

Table 1026 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1469	1624
R35137_PEA_1_PEA_1_PEA_1_T5	1402	1557
R35137_PEA_1_PEA_1_PEA_1_T10	1569	1724
R35137_PEA_1_PEA_1_PEA_1_T11	1502	1657
R35137_PEA_1_PEA_1_PEA_1_T12	1569	1724

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_27 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1027 below describes the starting and ending position of this segment on each transcript.

Table 1027 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1876	3898
R35137_PEA_1_PEA_1_PEA_1_T5	1752	3774
R35137_PEA_1_PEA_1_PEA_1_T10	1976	3998
R35137_PEA_1_PEA_1_PEA_1_T11	2049	4071
R35137_PEA_1_PEA_1_PEA_1_T12	2116	4138
R35137_PEA_1_PEA_1_PEA_1_T14	1134	1250

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_5 according to the present invention is supported by 20 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1028 below describes the starting and ending position of this segment on each transcript.

Table 1028 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	433	522
R35137_PEA_1_PEA_1_PEA_1_T5	433	522
R35137_PEA_1_PEA_1_PEA_1_T10	433	522
R35137_PEA_1_PEA_1_PEA_1_T11	433	522
R35137_PEA_1_PEA_1_PEA_1_T12	433	522
R35137_PEA_1_PEA_1_PEA_1_T14	433	522

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_7 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1029 below describes the starting and ending position of this segment on each transcript.

Table 1029 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	523	631

1050

R35137_PEA_1_PEA_1_PEA_1_T5	523	631
R35137_PEA_1_PEA_1_PEA_1_T10	523	631
R35137_PEA_1_PEA_1_PEA_1_T11	523	631
R35137_PEA_1_PEA_1_PEA_1_T12	523	631
R35137_PEA_1_PEA_1_PEA_1_T14	523	631

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_12 according to the present invention is supported by 22 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1030 below describes the starting and ending position of this segment on each transcript.

Table 1030 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	956	1009
R35137_PEA_1_PEA_1_PEA_1_T5	956	1009
R35137_PEA_1_PEA_1_PEA_1_T10	956	1009
R35137_PEA_1_PEA_1_PEA_1_T11	956	1009
R35137_PEA_1_PEA_1_PEA_1_T12	956	1009

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_14 according to the present invention is supported by 23 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and

R35137_PEA_1_PEA_1_PEA_1_T12. Table 1031 below describes the starting and ending position of this segment on each transcript.

Table 1031 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1010	1089
R35137_PEA_1_PEA_1_PEA_1_T5	1010	1089
R35137_PEA_1_PEA_1_PEA_1_T10	1010	1089
R35137_PEA_1_PEA_1_PEA_1_T11	1010	1089
R35137_PEA_1_PEA_1_PEA_1_T12	1010	1089

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_15 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T10 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1032 below describes the starting and ending position of this segment on each transcript.

Table 1032 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1090	1156
R35137_PEA_1_PEA_1_PEA_1_T10	1090	1156
R35137_PEA_1_PEA_1_PEA_1_T12	1090	1156

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_17 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and

R35137_PEA_1_PEA_1_PEA_1_T12. Table 1033 below describes the starting and ending position of this segment on each transcript.

Table 1033 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T10	1294	1393
R35137_PEA_1_PEA_1_PEA_1_T11	1227	1326
R35137_PEA_1_PEA_1_PEA_1_T12	1294	1393

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_21 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1034 below describes the starting and ending position of this segment on each transcript.

Table 1034 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T11	1658	1731
R35137_PEA_1_PEA_1_PEA_1_T12	1725	1798

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_22 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1035 below describes the starting and ending position of this segment on each transcript.

Table 1035 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1625	1697
R35137_PEA_1_PEA_1_PEA_1_T5	1558	1630
R35137_PEA_1_PEA_1_PEA_1_T10	1725	1797
R35137_PEA_1_PEA_1_PEA_1_T11	1732	1804
R35137_PEA_1_PEA_1_PEA_1_T12	1799	1871

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_23 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1036 below describes the starting and ending position of this segment on each transcript.

Table 1036 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1698	1737
R35137_PEA_1_PEA_1_PEA_1_T5	1631	1670
R35137_PEA_1_PEA_1_PEA_1_T10	1798	1837
R35137_PEA_1_PEA_1_PEA_1_T11	1805	1844
R35137_PEA_1_PEA_1_PEA_1_T12	1872	1911
R35137_PEA_1_PEA_1_PEA_1_T14	956	995

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_24 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously

described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T11 and R35137_PEA_1_PEA_1_PEA_1_T12. Table 1037 below describes the starting and ending position of this segment on each transcript.

Table 1037 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T11	1845	1910
R35137_PEA_1_PEA_1_PEA_1_T12	1912	1977

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_25 according to the present invention is supported by 30 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T5, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1038 below describes the starting and ending position of this segment on each transcript.

Table 1038 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1738	1818
R35137_PEA_1_PEA_1_PEA_1_T5	1671	1751
R35137_PEA_1_PEA_1_PEA_1_T10	1838	1918
R35137_PEA_1_PEA_1_PEA_1_T11	1911	1991
R35137_PEA_1_PEA_1_PEA_1_T12	1978	2058
R35137_PEA_1_PEA_1_PEA_1_T14	996	1076

Segment cluster R35137_PEA_1_PEA_1_PEA_1_node_26 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously

described. This segment can be found in the following transcript(s): R35137_PEA_1_PEA_1_PEA_1_T3, R35137_PEA_1_PEA_1_PEA_1_T10, R35137_PEA_1_PEA_1_PEA_1_T11, R35137_PEA_1_PEA_1_PEA_1_T12 and R35137_PEA_1_PEA_1_PEA_1_T14. Table 1039 below describes the starting and ending position of this segment on each transcript.

Table 1039 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R35137_PEA_1_PEA_1_PEA_1_T3	1819	1875
R35137_PEA_1_PEA_1_PEA_1_T10	1919	1975
R35137_PEA_1_PEA_1_PEA_1_T11	1992	2048
R35137_PEA_1_PEA_1_PEA_1_T12	2059	2115
R35137_PEA_1_PEA_1_PEA_1_T14	1077	1133

Variant protein alignment to the previously known protein:

```
Sequence name: ALAT_HUMAN_V1

Sequence documentation:

Alignment of: R35137_PEA_1_PEA_1_PEA_1_P9 x ALAT_HUMAN_V1    ..

Alignment segment 1/1:
```

1056

Quality: 2619.00

Escore: 0

Matching length: 274 Total length: 274

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
                .         .         .         .         .
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
                .         .         .         .         .
    101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                .         .         .         .         .
    151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                .         .         .         .         .
    201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
```

```
                    .          .
        251 TRECIEAVIRFAFEERLFLLADEV                                274
            ||||||||||||||||||||||||
        251 TRECIEAVIRFAFEERLFLLADEV                                274
```

Sequence name: ALAT_HUMAN_V1

Sequence documentation:

Alignment of: R35137_PEA_1_PEA_1_PEA_1_P8 x ALAT_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 3088.00
Escore: 0
Matching length: 320 Total length: 320
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                    .          .          .          .          .
          1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
```

1058

```
                                                                    
     1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ  50
                                                                    
    51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
                                                                    
    51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
                                                                    
   101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                                                                    
   101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                                                                    
   151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                                                                    
   151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                                                                    
   201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
                                                                    
   201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
                                                                    
   251 TRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPY 300
                                                                    
   251 TRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPY 300
                                                                    
   301 AGQQELASFHSTSKGYMGEC                                  320
                                                                    
   301 AGQQELASFHSTSKGYMGEC                                  320
```

Sequence name: ALAT_HUMAN_V1

Sequence documentation:

Alignment of: R35137_PEA_1_PEA_1_PEA_1_P11 x ALAT_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 2487.00
Escore: 0
Matching length: 271 Total length: 496
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 54.64 Total Percent Identity: 54.64
Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
                .         .         .         .         .
       51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
       51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
                .         .         .         .         .
      101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                .         .         .         .         .
151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                .         .         .         .         .
201 AVQVDYYLDEERAWALDVAELHRALGQAR..................... 229
    ||||||||||||||||||||||||||||
201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
                .         .         .         .         .
229 ..................................................... 229

251 TRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPY 300
                .         .         .         .         .
229 ..................................................... 229

301 AGQQELASFHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCP 350
                .         .         .         .         .
229 ..................................................... 229

351 PVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAELAAKAKLTEQVFNEA 400
                .         .         .         .         .
229 ..................................................... 229

401 PGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGIC 450
                .         .         .         .
230 ....SGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS     271
        ||||||||||||||||||||||||||||||||||||||||||
451 VVPGSGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS     496
```

Sequence name: ALAT_HUMAN_V1

Sequence documentation:

Alignment of: R35137_PEA_1_PEA_1_PEA_1_P2 x ALAT_HUMAN_V1 ..

Alignment segment 1/1:

Quality: 2619.00
Escore: 0
Matching length: 274 Total length: 274
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
                .         .         .         .         .
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
```

```
                .         .         .         .         .
      101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                .         .         .         .         .
      151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                .         .         .         .         .
      201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
      201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
                .         .
      251 TRECIEAVIRFAFEERLFLLADEV                           274
          ||||||||||||||||||||||||
      251 TRECIEAVIRFAFEERLFLLADEV                           274
```

Sequence name: ALAT_HUMAN_V1

Sequence documentation:

Alignment of: R35137_PEA_1_PEA_1_PEA_1_P4 x ALAT_HUMAN_V1 ..

Alignment segment 1/1:

1063

Quality: 4785.00

Escore: 0

Matching length: 494 Total length: 494

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQ 50
                .         .         .         .         .
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ELRQGVKKPFTEVIRANIGDAQAMGQRPITFLRQVLALCVNPDLLSSPNF 100
                .         .         .         .         .
    101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 PDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPAD 150
                .         .         .         .         .
    151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 PNNVFLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELG 200
                .         .         .         .         .
    201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 AVQVDYYLDEERAWALDVAELHRALGQARDHCRPRALCVINPGNPTGQVQ 250
```

```
                .         .         .         .         .
251 TRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPY 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 TRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPY 300
                .         .         .         .         .
301 AGQQELASFHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCP 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 AGQQELASFHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCP 350
                .         .         .         .         .
351 PVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAELAAKAKLTEQVFNEA 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 PVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAELAAKAKLTEQVFNEA 400
                .         .         .         .         .
401 PGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGIC 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 PGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGIC 450
                .         .         .         .
451 VVPGSGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLE       494
    ||||||||||||||||||||||||||||||||||||||||||||
451 VVPGSGFGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLE       494
```

DESCRIPTION FOR CLUSTER Z25299

Cluster Z25299 features 5 transcript(s) and 11 segment(s) of interest, the names for which are given in Tables 1040 and 1041, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1042.

Table 1040 - Transcripts of interest

Transcript Name	Sequence ID No.
Z25299_PEA_2_T1	120
Z25299_PEA_2_T2	121
Z25299_PEA_2_T3	122

Z25299_PEA_2_T6	123
Z25299_PEA_2_T9	124

Table 1041 - Segments of interest

Segment Name	Sequence ID No.
Z25299_PEA_2_node_20	876
Z25299_PEA_2_node_21	877
Z25299_PEA_2_node_23	878
Z25299_PEA_2_node_24	879
Z25299_PEA_2_node_8	880
Z25299_PEA_2_node_12	881
Z25299_PEA_2_node_13	882
Z25299_PEA_2_node_14	883
Z25299_PEA_2_node_17	884
Z25299_PEA_2_node_18	885
Z25299_PEA_2_node_19	886

Table 1042 - Proteins of interest

Protein Name	Sequence ID No.
Z25299_PEA_2_P2	1390
Z25299_PEA_2_P3	1391
Z25299_PEA_2_P7	1392
Z25299_PEA_2_P10	1393

These sequences are variants of the known protein Antileukoproteinase 1 precursor (SwissProt accession identifier ALK1_HUMAN; known also according to the synonyms ALP; HUSI-1; Seminal proteinase inhibitor; Secretory leukocyte protease inhibitor; BLPI; Mucus proteinase inhibitor; MPI; WAP four-disulfide core domain protein 4; Protease inhibitor WAP4), SEQ ID NO: 1454, referred to herein as the previously known protein.

Protein Antileukoproteinase 1 precursor is known or believed to have the following function(s): Acid-stable proteinase inhibitor with strong affinities for trypsin, chymotrypsin, elastase, and cathepsin G. May prevent elastase-mediated damage to oral and possibly other mucosal tissues. The sequence for protein Antileukoproteinase 1 precursor is given at the end of the application, as "Antileukoproteinase 1 precursor amino acid sequence". Protein Antileukoproteinase 1 precursor localization is believed to be Secreted.

It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Elastase inhibitor; Tryptase inhibitor. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Anti-inflammatory; Antiasthma.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: proteinase inhibitor; serine protease inhibitor, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster Z25299 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 35 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 35 and Table 1043. This cluster is overexpressed (at least at a minimum level) in the

following pathological conditions: brain malignant tumors, a mixture of malignant tumors from different tissues and ovarian carcinoma.

Table 1043 - Normal tissue distribution

Name of Tissue	Number
bladder	82
bone	6
brain	0
colon	37
epithelial	145
general	73
head and neck	638
kidney	26
liver	68
lung	465
breast	52
ovary	0
pancreas	20
prostate	36
skin	215
stomach	219
uterus	113

5

Table 1044 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bladder	8.2e-01	8.5e-01	9.2e-01	0.6	9.7e-01	0.5
bone	5.5e-01	7.3e-01	4.0e-01	2.1	4.9e-01	1.5
brain	8.8e-02	1.5e-01	2.3e-03	7.7	1.2e-02	4.8
colon	3.3e-01	2.8e-01	4.2e-01	1.6	4.2e-01	1.5

epithelial	2.5e-01	7.6e-01	3.8e-01	1.0	1	0.6
general	6.4e-03	2.5e-01	1.7e-06	1.6	5.2e-01	0.9
head and neck	3.6e-01	5.9e-01	7.6e-01	0.6	1	0.3
kidney	7.4e-01	8.4e-01	2.1e-01	2.1	4.2e-01	1.4
liver	4.1e-01	9.1e-01	4.2e-02	3.2	6.4e-01	0.8
lung	7.6e-01	8.3e-01	9.8e-01	0.5	1	0.3
breast	5.0e-01	5.5e-01	9.8e-02	1.6	3.4e-01	1.1
ovary	3.7e-02	3.0e-02	6.9e-03	6.1	4.9e-03	5.6
pancreas	3.8e-01	3.6e-01	3.6e-01	1.7	3.9e-01	1.5
prostate	9.1e-01	9.2e-01	8.9e-01	0.5	9.4e-01	0.5
skin	6.0e-01	8.1e-01	9.3e-01	0.4	1	0.1
stomach	3.0e-01	8.1e-01	9.1e-01	0.6	1	0.3
uterus	1.6e-01	1.3e-01	3.2e-02	1.6	3.0e-01	1.1

As noted above, cluster Z25299 features 5 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Antileukoproteinase 1 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein Z25299_PEA_2_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z25299_PEA_2_T1. An alignment is given to the known protein (Antileukoproteinase 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z25299_PEA_2_P2 and ALK1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z25299_PEA_2_P2, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLK CCMGMCGKSCVSPVK corresponding to amino acids 1 - 131 of ALK1_HUMAN, which also

corresponds to amino acids 1 - 131 of Z25299_PEA_2_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GKQGMRAH corresponding to amino acids 132 - 139 of Z25299_PEA_2_P2, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z25299_PEA_2_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GKQGMRAH in Z25299_PEA_2_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z25299_PEA_2_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1045, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table1045 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
136	M -> T	Yes
20	P ->	No
43	C -> R	No
48	K -> N	No
83	R -> K	No

84	R -> W	No

Variant protein Z25299_PEA_2_P2 is encoded by the following transcript(s): Z25299_PEA_2_T1, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z25299_PEA_2_T1 is shown in bold; this coding portion starts at position 124 and ends at position 540. The transcript also has the following SNPs as listed in Table 1046 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1046 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
122	C -> T	No
123	C -> T	No
530	T -> C	Yes
989	C -> T	Yes
1127	C -> T	Yes
1162	A -> C	Yes
1180	A -> C	Yes
1183	A -> C	Yes
1216	A -> C	Yes
1262	G -> A	Yes
183	T ->	No
250	T -> C	No
267	A -> C	No
267	A -> G	No
339	C -> T	Yes
371	G -> A	No
373	A -> T	No

435	C -> T	No

Variant protein Z25299_PEA_2_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z25299_PEA_2_T2. An alignment is given to the known protein (Antileukoproteinase 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z25299_PEA_2_P3 and ALK1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z25299_PEA_2_P3, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLK CCMGMCGKSCVSPVK corresponding to amino acids 1 - 131 of ALK1_HUMAN, which also corresponds to amino acids 1 - 131 of Z25299_PEA_2_P3, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GEKRHHKQLRDQEVDPLEMRRHSAG corresponding to amino acids 132 - 156 of Z25299_PEA_2_P3, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z25299_PEA_2_P3, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GEKRHHKQLRDQEVDPLEMRRHSAG in Z25299_PEA_2_P3.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide

prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z25299_PEA_2_P3 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1047, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1047 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
20	P ->	No
43	C -> R	No
48	K -> N	No
83	R -> K	No
84	R -> W	No

Variant protein Z25299_PEA_2_P3 is encoded by the following transcript(s): Z25299_PEA_2_T2, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z25299_PEA_2_T2 is shown in bold; this coding portion starts at position 124 and ends at position 591. The transcript also has the following SNPs as listed in Table 1048 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1048 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
122	C -> T	No

123	C -> T	No
183	T ->	No
250	T -> C	No
267	A -> C	No
267	A -> G	No
339	C -> T	Yes
371	G -> A	No
373	A -> T	No
435	C -> T	No

Variant protein Z25299_PEA_2_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z25299_PEA_2_T6. An alignment is given to the known protein (Antileukoproteinase 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z25299_PEA_2_P7 and ALK1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z25299_PEA_2_P7, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNP corresponding to amino acids 1 - 81 of ALK1_HUMAN, which also corresponds to amino acids 1 - 81 of Z25299_PEA_2_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RGSLGSAQ corresponding to amino acids 82 - 89 of Z25299_PEA_2_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of Z25299_PEA_2_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RGSLGSAQ in Z25299_PEA_2_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z25299_PEA_2_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1049, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1049 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
20	P ->	No
43	C -> R	No
48	K -> N	No
82	R -> S	No

Variant protein Z25299_PEA_2_P7 is encoded by the following transcript(s): Z25299_PEA_2_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z25299_PEA_2_T6 is shown in bold; this coding portion starts at position 124 and ends at position 390. The transcript also has the following SNPs as listed in Table 1050 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein Z25299_PEA_2_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1050 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
122	C -> T	No
123	C -> T	No
576	A -> C	Yes
594	A -> C	Yes
597	A -> C	Yes
630	A -> C	Yes
676	G -> A	Yes
183	T ->	No
250	T -> C	No
267	A -> C	No
267	A -> G	No
339	C -> T	Yes
369	A -> T	No
431	C -> T	No
541	C -> T	Yes

Variant protein Z25299_PEA_2_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) Z25299_PEA_2_T9. An alignment is given to the known protein (Antileukoproteinase 1 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between Z25299_PEA_2_P10 and ALK1_HUMAN:

1.An isolated chimeric polypeptide encoding for Z25299_PEA_2_P10, comprising a first amino acid sequence being at least 90 % homologous to MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCP GKKRCCPDTCGIKCLDPVDTPNPT corresponding to amino acids 1 - 82 of ALK1_HUMAN, which also corresponds to amino acids 1 - 82 of Z25299_PEA_2_P10.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein Z25299_PEA_2_P10 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1051, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1051 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
20	P ->	No
43	C -> R	No
48	K -> N	No

Variant protein Z25299_PEA_2_P10 is encoded by the following transcript(s): Z25299_PEA_2_T9, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript Z25299_PEA_2_T9 is shown in bold; this coding portion starts at position 124 and ends at position 369. The transcript also has the following SNPs as listed in Table 1052 (given according to their position on the nucleotide sequence, with the alternative

nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein Z25299_PEA_2_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1052 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
122	C -> T	No
123	C -> T	No
451	A -> C	Yes
484	A -> C	Yes
530	G -> A	Yes
183	T ->	No
250	T -> C	No
267	A -> C	No
267	A -> G	No
339	C -> T	Yes
395	C -> T	Yes
430	A -> C	Yes
448	A -> C	Yes

As noted above, cluster Z25299 features 11 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster Z25299_PEA_2_node_20 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1. Table 1053 below describes the starting and ending position of this segment on each transcript.

Table 1053 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	518	1099

Segment cluster Z25299_PEA_2_node_21 according to the present invention is supported by 162 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T6 and Z25299_PEA_2_T9. Table 1054 below describes the starting and ending position of this segment on each transcript.

Table 1054 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	1100	1292
Z25299_PEA_2_T6	514	706
Z25299_PEA_2_T9	368	560

Segment cluster Z25299_PEA_2_node_23 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T2. Table 1055 below describes the starting and ending position of this segment on each transcript.

Table 1055 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T2	518	707

Segment cluster Z25299_PEA_2_node_24 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T2 and Z25299_PEA_2_T3. Table 1056 below describes the starting and ending position of this segment on each transcript.

Table 1056 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T2	708	886
Z25299_PEA_2_T3	518	696

Segment cluster Z25299_PEA_2_node_8 according to the present invention is supported by 218 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3, Z25299_PEA_2_T6 and Z25299_PEA_2_T9. Table 1057 below describes the starting and ending position of this segment on each transcript.

Table 1057 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	1	208
Z25299_PEA_2_T2	1	208
Z25299_PEA_2_T3	1	208
Z25299_PEA_2_T6	1	208
Z25299_PEA_2_T9	1	208

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster Z25299_PEA_2_node_12 according to the present invention is supported by 228 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3, Z25299_PEA_2_T6 and Z25299_PEA_2_T9. Table 1058 below describes the starting and ending position of this segment on each transcript.

Table 1058 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	209	245
Z25299_PEA_2_T2	209	245
Z25299_PEA_2_T3	209	245
Z25299_PEA_2_T6	209	245
Z25299_PEA_2_T9	209	245

Segment cluster Z25299_PEA_2_node_13 according to the present invention is supported by 246 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3, Z25299_PEA_2_T6 and Z25299_PEA_2_T9. Table 1059 below describes the starting and ending position of this segment on each transcript.

Table 1059 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	246	357
Z25299_PEA_2_T2	246	357
Z25299_PEA_2_T3	246	357
Z25299_PEA_2_T6	246	357
Z25299_PEA_2_T9	246	357

Segment cluster Z25299_PEA_2_node_14 according to the present invention can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3, Z25299_PEA_2_T6 and Z25299_PEA_2_T9. Table 1060 below describes the starting and ending position of this segment on each transcript.

Table 1060 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	358	367
Z25299_PEA_2_T2	358	367

Z25299_PEA_2_T3	358	367
Z25299_PEA_2_T6	358	367
Z25299_PEA_2_T9	358	367

Segment cluster Z25299_PEA_2_node_17 according to the present invention can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2 and Z25299_PEA_2_T3. Table 1061 below describes the starting and ending position of this segment on each transcript.

Table 1061 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	368	371
Z25299_PEA_2_T2	368	371
Z25299_PEA_2_T3	368	371

Segment cluster Z25299_PEA_2_node_18 according to the present invention is supported by 221 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3 and Z25299_PEA_2_T6. Table 1062 below describes the starting and ending position of this segment on each transcript.

Table 1062 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	372	427
Z25299_PEA_2_T2	372	427
Z25299_PEA_2_T3	372	427
Z25299_PEA_2_T6	368	423

Segment cluster Z25299_PEA_2_node_19 according to the present invention is supported by 197 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): Z25299_PEA_2_T1, Z25299_PEA_2_T2, Z25299_PEA_2_T3 and Z25299_PEA_2_T6. Table 1063 below describes the starting and ending position of this segment on each transcript.

Table 1063 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
Z25299_PEA_2_T1	428	517
Z25299_PEA_2_T2	428	517
Z25299_PEA_2_T3	428	517
Z25299_PEA_2_T6	424	513

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/oXgeQ4MeyL/K6Vqb1MQu2:ALK1_HUMAN

Sequence documentation:

Alignment of: Z25299_PEA_2_P2 x ALK1_HUMAN    ..

Alignment segment 1/1:

                    Quality: 1371.00
Escore:        0
              Matching length:      131                 Total
length:      131
```

1083

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
                .         .         .         .         .
     51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLN 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLN 100
                .         .         .
    101 PPNFCEMDGQCKRDLKCCMGMCGKSCVSPVK                    131
        ||||||||||||||||||||||||||||||
    101 PPNFCEMDGQCKRDLKCCMGMCGKSCVSPVK                    131
```

Sequence name: /tmp/rbf314VLIm/yR43i4SbP4:ALK1_HUMAN

Sequence documentation:

Alignment of: Z25299_PEA_2_P3 x ALK1_HUMAN ..

Alignment segment 1/1:

Quality: 1371.00

Escore: 0

Matching length: 131 Total length: 131

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
                .         .         .         .         .
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLN 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPTRRKPGKCPVTYGQCLMLN 100
                .         .         .
      101 PPNFCEMDGQCKRDLKCCMGMCGKSCVSPVK                    131
          ||||||||||||||||||||||||||||||
      101 PPNFCEMDGQCKRDLKCCMGMCGKSCVSPVK                    131
```

Sequence name: /tmp/KCtSXACZXe/rK4T6LKeRX:ALK1_HUMAN

Sequence documentation:

Alignment of: Z25299_PEA_2_P7 x ALK1_HUMAN ..

Alignment segment 1/1:

Quality: 835.00
Escore: 0
Matching length: 81 Total length: 81
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                    .         .         .         .         .
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
                  .         .         .
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNP                    81
          ||||||||||||||||||||||||||||||||
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNP                    81
```

```
Sequence name: /tmp/LcBlcAxB6c/NSI9pqfxoU:ALK1_HUMAN

Sequence documentation:

Alignment of: Z25299_PEA_2_P10 x ALK1_HUMAN    ..

Alignment segment 1/1:

                      Quality:  844.00
Escore:       0
             Matching length:      82                Total
length:      82
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                         Gaps:       0

Alignment:
                  .         .         .         .         .
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPE 50
                  .         .         .
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPT                    82
          ||||||||||||||||||||||||||||||||
       51 CQSDWQCPGKKRCCPDTCGIKCLDPVDTPNPT                    82
```

Expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitorZ25299 transcripts, which are detectable by amplicon as depicted in sequence name Z25299 junc13-14-21 in normal and cancerous lung tissues

Expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by or according to junc13-14-21, Z25299 junc13-14-21 amplicon (SEQ ID NO: 1666) and Z25299 junc13-14-21F (SEQ ID NO: 1664) and Z25299 junc13-14-21R (SEQ ID NO: 1665) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2 "Tissue sample in testing panel", above), to obtain a value of fold differential expression for each sample relative to median of the normal PM samples.

Figure 36 is a histogram showing down regulation of the above-indicated Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 36, the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by the above amplicon(s) in cancer samples was significantly lower than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue sample in testing panel").

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by the above amplicon(s) in lung cancer samples versus the normal tissue samples was determined by T test as 1.98E-04. This value demonstrates statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Z25299 junc13-14-21F forward primer; and Z25299 junc13-14-21R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Z25299 junc13-14-21.

Forward primer (SEQ ID NO: 1664): ACCCCAAACCCAACTTGATTC

Reverse primer (SEQ ID NO: 1665): TCAGTGGTGGAGCCAAGTCTC

Amplicon (SEQ ID NO: 1666):

ACCCCAAACCCAACTTGATTCCTGCCATATGGAGGAGGCTCTGGAGTCCTGCTCTGT GTGGTCCAGGTCCTTTCCACCCTGAGACTTGGCTCCACCACTGA

Z25299 transcripts, which are detectable by amplicon as depicted in sequence name Z25299 seg20 in normal and cancerous lung tissues

Expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by or according to seg20, Z25299 seg20 amplicon (SEQ ID NO: 1669) and Z25299 seg20F (SEQ ID NO: 1667) and Z25299 seg20R (SEQ ID NO: 1668) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above). Then the reciprocal of this ratio was calculated, to obtain a value of fold down-regulation for each sample relative to median of the normal PM samples.

Figure 37 is a histogram showing down regulation of the above-indicated Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 5 fold down regulation, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 37, the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by the above amplicon(s) in cancer samples was significantly lower than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue sample in testing panel"). Notably an down regulation of at least 5 fold was found in 6 out of 15 adenocarcinoma samples, 9 out of 16 squamous cell carcinoma samples, 3 out of 4 large cell carcinoma samples and in 8 out of 8 small cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by the above amplicon(s) in lung cancer samples versus the normal tissue samples was determined by T test as 9.43E-02 in adenocarcinoma, 5.62E-02 in squamous cell carcinoma, 3.38E-01 in large cell carcinoma and 3.78E-02 in small cell carcinoma.

Threshold of 5 fold down regulation was found to differentiate between cancer and normal samples with P value of 3.73E-02 in adenocarcinoma, 1.10E-02 in squamous cell carcinoma, 2.64E-02 in large cell carcinoma and 7.14E-05 in small cell carcinoma checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Z25299 seg20F forward primer; and Z25299 seg20R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Z25299 seg20.

Forward primer (SEQ ID NO: 1667): CTCCTGAACCCTACTCCAAGCA

Reverse primer (SEQ ID NO: 1668): CAGGCGATCCTATGGAAATCC

Amplicon (SEQ ID NO: 1669):

CTCCTGAACCCTACTCCAAGCACAGCCTCTGTCTGACTCCCTTGTCCTTCAAGAGAA CTGTTCTCCAGGTCTCAGGGCCAGGATTTCCATAGGATCGCCTG

Expression of Homo sapiens secretory leukocyte protease inhibitor (antileukoproteinase) (SLPI) *Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299 seg23 in normal and cancerous lung tissues*

Expression of Homo sapiens secretory leukocyte protease inhibitor (antileukoproteinase) (SLPI) transcripts detectable by or according to seg23, Z25299 seg23 amplicon (SEQ ID NO: 1672) and primers Z25299 seg23F (SEQ ID NO: 1670) and Z25299 seg23R (SEQ ID NO: 1671) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above). Then the reciprocal of this ratio was calculated, to obtain a value of fold down-regulation for each sample relative to median of the normal PM samples.

Figure 68 is a histogram showing down regulation of the above-indicated Homo sapiens secretory leukocyte protease inhibitor (antileukoproteinase) (SLPI) transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 68, the expression of Homo sapiens secretory leukocyte protease inhibitor (antileukoproteinase) (SLPI) transcripts detectable by the above amplicon(s) in cancer samples was significantly lower than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2). Notably down regulation of at least 10 fold was found in 7 out of 15 adenocarcinoma samples, 9 out of 16 squamous cell carcinoma samples, 3 out of 4 large cell carcinoma samples and in 8 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Z25299 seg23F forward primer; and Z25299 seg23R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Z25299 seg23.

Primers:

Forward primer Z25299 seg23F (SEQ ID NO: 1670): CAAGCAATTGAGGGACCAGG

Reverse primer Z25299 seg23R (SEQ ID NO: 1671):

CAAAAAACATTGTTAATGAGAGAGATGAC

Amplicon Z25299 seg23F (SEQ ID NO: 1672):

CAAGCAATTGAGGGACCAGGAAGTGGATCCTCTAGAGATGAGGAGGCATTCTGCTG GATGACTTTTAAAAATGTTTTCTCCAGAGTCATCTCTCTCATTAACAATGTTTTTTG

Expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg20 in different normal tissues

Expression of Secretory leukocyte protease inhibitor transcripts detectable by or according to *Z25299seg20* amplicon (SEQ ID NO: 1669) and primers: *Z25299seg23F* (SEQ ID NO: 1667) *Z25299seg20R* (SEQ ID NO: 1668) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20, Table 3), to obtain a value of relative expression of each sample relative to median of the ovary samples.

Primers:

Forward primer (SEQ ID NO: 1667): CTCCTGAACCCTACTCCAAGCA

Reverse primer (SEQ ID NO: 1668): CAGGCGATCCTATGGAAATCC

Amplicon (SEQ ID NO: 1669):

CTCCTGAACCCTACTCCAAGCACAGCCTCTGTCTGACTCCCTTGTCCTTCAAGAGAA CTGTTCTCCAGGTCTCAGGGCCAGGATTTCCATAGGATCGCCTG

The results are demonstrated in Figure 69, showing the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg20 in different normal tissues.

Expression of Secretory leukocyte protease inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg23 in different normal tissues

Expression of Secretory leukocyte protease inhibitor transcripts detectable by or according to *Z25299seg23* amplicon (SEQ ID NO: 1672) and primers: *Z25299seg23F* (SEQ ID NO: 1670) *Z25299seg23R* (SEQ ID NO: 1671) was measured by real time PCR. In parallel the expression of four housekeeping genes –RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20, Table 3), to obtain a value of relative expression of each sample relative to median of the ovary samples.

Primers:

Forward primer Z25299 seg23F (SEQ ID NO: 1670): CAAGCAATTGAGGGACCAGG

Reverse primer Z25299 seg23R (SEQ ID NO: 1671):
CAAAAAACATTGTTAATGAGAGAGATGAC

Amplicon Z25299 seg23F (SEQ ID NO: 1672):
CAAGCAATTGAGGGACCAGGAAGTGGATCCTCTAGAGATGAGGAGGCATTCTGCTG
GATGACTTTTAAAAATGTTTTCTCCAGAGTCATCTCTCTCATTAACAATGTTTTTTG

The results are demonstrated in Figure 70, showing the expression of Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor Z25299 transcripts which are detectable by amplicon as depicted in sequence name Z25299seg23 in different normal tissues.

DESCRIPTION FOR CLUSTER HSSTROL3

Cluster HSSTROL3 features 6 transcript(s) and 16 segment(s) of interest, the names for which are given in Tables 1064 and 1065, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1066.

Table 1064 - Transcripts of interest

Transcript Name	Sequence ID No.
HSSTROL3_T5	125
HSSTROL3_T8	126
HSSTROL3_T9	127
HSSTROL3_T10	128
HSSTROL3_T11	129
HSSTROL3_T12	130

Table 1065 - Segments of interest

Segment Name	Sequence ID No.
HSSTROL3_node_6	887
HSSTROL3_node_10	888
HSSTROL3_node_13	889
HSSTROL3_node_15	890
HSSTROL3_node_19	891
HSSTROL3_node_21	892
HSSTROL3_node_24	893
HSSTROL3_node_25	894
HSSTROL3_node_26	895
HSSTROL3_node_28	896
HSSTROL3_node_29	897
HSSTROL3_node_11	898
HSSTROL3_node_17	899
HSSTROL3_node_18	900

HSSTROL3_node_20	901
HSSTROL3_node_27	902

Table 1066 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HSSTROL3_P4	1394	HSSTROL3_T5
HSSTROL3_P5	1395	HSSTROL3_T8; HSSTROL3_T9
HSSTROL3_P7	1396	HSSTROL3_T10
HSSTROL3_P8	1397	HSSTROL3_T11
HSSTROL3_P9	1398	HSSTROL3_T12

These sequences are variants of the known protein Stromelysin-3 precursor (SwissProt accession identifier MM11_HUMAN; known also according to the synonyms EC 3.4.24.-; Matrix metalloproteinase-11; MMP-11; ST3; SL-3), SEQ ID NO: 1455, referred to herein as the previously known protein.

Protein Stromelysin-3 precursor is known or believed to have the following function(s): May play an important role in the progression of epithelial malignancies. The sequence for protein Stromelysin-3 precursor is given at the end of the application, as "Stromelysin-3 precursor amino acid sequence".

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: proteolysis and peptidolysis; developmental processes; morphogenesis, which are annotation(s) related to Biological Process; stromelysin 3; calcium binding; zinc binding; hydrolase, which are annotation(s) related to Molecular Function; and extracellular matrix, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HSSTROL3 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the left hand column of the table and the numbers on the y-axis of figure 38 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 38 and Table 1067. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: transitional cell carcinoma, epithelial malignant tumors, a mixture of malignant tumors from different tissues and pancreas carcinoma.

Table 1067 - Normal tissue distribution

Name of Tissue	Number
adrenal	0
bladder	0
brain	1
colon	63
epithelial	33
general	13
head and neck	101
kidney	0
lung	11
breast	8
ovary	14
pancreas	0
prostate	2
skin	99
Thyroid	0
uterus	181

Table 1068 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	1	4.6e-01	1	1.0	5.3e-01	1.9
bladder	2.7e-01	3.4e-01	3.3e-03	4.9	2.1e-02	3.3
brain	3.5e-01	2.6e-01	1	1.7	3.3e-01	2.8
colon	7.7e-02	1.5e-01	3.1e-01	1.4	5.2e-01	1.0
epithelial	1.2e-04	1.2e-02	1.3e-06	2.7	4.6e-02	1.4
general	5.4e-09	3.1e-05	1.8e-16	5.0	3.1e-07	2.6
head and neck	4.6e-01	4.3e-01	1	0.6	9.4e-01	0.7
kidney	2.5e-01	3.5e-01	1.1e-01	4.0	2.4e-01	2.8
lung	1.8e-01	4.5e-01	1.9e-01	2.7	5.1e-01	1.4
breast	2.0e-01	3.4e-01	7.3e-02	3.3	2.5e-01	2.0
ovary	2.6e-01	3.2e-01	2.2e-02	2.0	7.0e-02	1.6
pancreas	9.5e-02	1.8e-01	1.8e-04	7.8	1.6e-03	5.5
prostate	8.2e-01	7.8e-01	4.5e-01	1.8	5.6e-01	1.5
skin	5.2e-01	5.8e-01	7.1e-01	0.8	1	0.3
Thyroid	2.9e-01	2.9e-01	1	1.1	1	1.1
uterus	4.2e-01	8.0e-01	7.5e-01	0.6	9.9e-01	0.4

As noted above, cluster HSSTROL3 features 6 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Stromelysin-3 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HSSTROL3_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSSTROL3_T5. An alignment is given to the known protein (Stromelysin-3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSSTROL3_P4 and MM11_HUMAN:

1.An isolated chimeric polypeptide encoding for HSSTROL3_P4, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 ofMM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P4, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P4, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRFPVHAALVWGPE KNKIYFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADG corresponding to amino acids 165 - 445 of MM11_HUMAN, which also corresponds to amino acids 165 - 445 of HSSTROL3_P4, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ALGVRQLVGGGHSSRFSHLVVAGLPHACHRKSGSSSQVLCPEPSALLSVAG corresponding to amino acids 446 - 496 of HSSTROL3_P4, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HSSTROL3_P4, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ALGVRQLVGGGHSSRFSHLVVAGLPHACHRKSGSSSQVLCPEPSALLSVAG in HSSTROL3_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSSTROL3_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1069, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1069 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
38	V -> A	Yes
104	R -> P	Yes
214	A ->	No
323	Q -> H	Yes

Variant protein HSSTROL3_P4 is encoded by the following transcript(s): HSSTROL3_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSSTROL3_T5 is shown in bold; this coding portion starts at position 24 and ends at position 1511. The transcript also has the following SNPs as listed in Table 1070 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1070 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes

334	G -> C	Yes
663	G ->	No
699	-> T	No
992	G -> C	Yes
1528	A -> G	Yes
1710	A -> G	Yes
2251	A -> G	Yes
2392	C ->	No
2444	C -> A	Yes
2470	A -> T	Yes
2687	-> G	No
2696	-> G	No
2710	C ->	No
2729	-> A	No
2755	T -> C	No
2813	A ->	No
2813	A -> C	No
2963	A ->	No
2963	A -> C	No
2993	T -> C	Yes
3140	-> T	No

Variant protein HSSTROL3_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSSTROL3_T8 and HSSTROL3_T9. An alignment is given to the known protein (Stromelysin-3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSSTROL3_P5 and MM11_HUMAN:

1.An isolated chimeric polypeptide encoding for HSSTROL3_P5, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P5, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P5, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQ corresponding to amino acids 165 - 358 of MM11_HUMAN, which also corresponds to amino acids 165 - 358 of HSSTROL3_P5, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence ELGFPSSTGRDESLEHCRCQGLHK corresponding to amino acids 359 - 382 of HSSTROL3_P5, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HSSTROL3_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence ELGFPSSTGRDESLEHCRCQGLHK in HSSTROL3_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSSTROL3_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1071, (given according to their position(s) on the

amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1071 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
38	V -> A	Yes
104	R -> P	Yes
214	A ->	No
323	Q -> H	Yes

Variant protein HSSTROL3_P5 is encoded by the following transcript(s): HSSTROL3_T8 and HSSTROL3_T9, for which the sequence(s) is/are given at the end of the application.

The coding portion of transcript HSSTROL3_T8 is shown in bold; this coding portion starts at position 24 and ends at position 1169. The transcript also has the following SNPs as listed in Table 1072 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1072 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes
334	G -> C	Yes
663	G ->	No
699	-> T	No
992	G -> C	Yes

1903	C ->	No
1955	C -> A	Yes
1981	A -> T	Yes
2198	-> G	No
2207	-> G	No
2221	C ->	No
2240	-> A	No
2266	T -> C	No
2324	A ->	No
2324	A -> C	No
2474	A ->	No
2474	A -> C	No
2504	T -> C	Yes
2651	-> T	No

The coding portion of transcript HSSTROL3_T9 is shown in bold; this coding portion starts at position 24 and ends at position 1169. The transcript also has the following SNPs as listed in Table 1073 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1073 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes
334	G -> C	Yes
663	G ->	No
699	-> T	No
992	G -> C	Yes
1666	A -> G	Yes

1848	A -> G	Yes
2389	A -> G	Yes
2530	C ->	No
2582	C -> A	Yes
2608	A -> T	Yes
2825	-> G	No
2834	-> G	No
2848	C ->	No
2867	-> A	No
2893	T -> C	No
2951	A ->	No
2951	A -> C	No
3101	A ->	No
3101	A -> C	No
3131	T -> C	Yes
3278	-> T	No

Variant protein HSSTROL3_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSSTROL3_T10. An alignment is given to the known protein (Stromelysin-3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSSTROL3_P7 and MM11_HUMAN:

1.An isolated chimeric polypeptide encoding for HSSTROL3_P7, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of

HSSTROL3_P7, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P7, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQG corresponding to amino acids 165 - 359 of MM11_HUMAN, which also corresponds to amino acids 165 - 359 of HSSTROL3_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TTGVSTPAPGV corresponding to amino acids 360 - 370 of HSSTROL3_P7, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HSSTROL3_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TTGVSTPAPGV in HSSTROL3_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSSTROL3_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1074, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1074 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
38	V -> A	Yes
104	R -> P	Yes
214	A ->	No
323	Q -> H	Yes

Variant protein HSSTROL3_P7 is encoded by the following transcript(s): HSSTROL3_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSSTROL3_T10 is shown in bold; this coding portion starts at position 24 and ends at position 1133. The transcript also has the following SNPs as listed in Table 1075 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1075 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes
334	G -> C	Yes
663	G ->	No
699	-> T	No
992	G -> C	Yes
1386	A -> G	Yes
1568	A -> G	Yes
2109	A -> G	Yes
2250	C ->	No
2302	C -> A	Yes
2328	A -> T	Yes

2545	-> G	No
2554	-> G	No
2568	C ->	No
2587	-> A	No
2613	T -> C	No
2671	A ->	No
2671	A -> C	No
2821	A ->	No
2821	A -> C	No
2851	T -> C	Yes
2998	-> T	No

Variant protein HSSTROL3_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSSTROL3_T11. An alignment is given to the known protein (Stromelysin-3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSSTROL3_P8 and MM11_HUMAN:

1.An isolated chimeric polypeptide encoding for HSSTROL3_P8, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFP WQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 1 - 163 of MM11_HUMAN, which also corresponds to amino acids 1 - 163 of HSSTROL3_P8, a bridging amino acid H corresponding to amino acid 164 of HSSTROL3_P8, a second amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLE corresponding to amino acids 165 - 286 of MM11_HUMAN, which also corresponds

to amino acids 165 - 286 of HSSTROL3_P8, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VRPCLPVPLLLCWPL corresponding to amino acids 287 - 301 of HSSTROL3_P8, wherein said first amino acid sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HSSTROL3_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VRPCLPVPLLLCWPL in HSSTROL3_P8.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSSTROL3_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1076, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1076 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
38	V -> A	Yes
104	R -> P	Yes
214	A ->	No

Variant protein HSSTROL3_P8 is encoded by the following transcript(s): HSSTROL3_T11, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSSTROL3_T11 is shown in bold; this coding portion starts at position 24 and ends at position 926. The transcript also has the following SNPs as listed in Table 1077 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1077 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes
334	G -> C	Yes
663	G ->	No
699	-> T	No
935	G -> A	Yes
948	G -> A	Yes
1084	G -> C	Yes
1557	C ->	No
1609	C -> A	Yes
1635	A -> T	Yes
1852	-> G	No
1861	-> G	No
1875	C ->	No
1894	-> A	No
1920	T -> C	No
1978	A ->	No
1978	A -> C	No
2128	A ->	No
2128	A -> C	No

2158	T -> C	Yes
2305	-> T	No

Variant protein HSSTROL3_P9 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSSTROL3_T12. An alignment is given to the known protein (Stromelysin-3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSSTROL3_P9 and MM11_HUMAN:

1.An isolated chimeric polypeptide encoding for HSSTROL3_P9, comprising a first amino acid sequence being at least 90 % homologous to MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSS PAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQK corresponding to amino acids 1 - 96 of MM11_HUMAN, which also corresponds to amino acids 1 - 96 of HSSTROL3_P9, a second amino acid sequence being at least 90 % homologous to RILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYW corresponding to amino acids 113 - 163 of MM11_HUMAN, which also corresponds to amino acids 97 - 147 of HSSTROL3_P9, a bridging amino acid H corresponding to amino acid 148 of HSSTROL3_P9, a third amino acid sequence being at least 90 % homologous to GDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLG LQHTTAAKALMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTN EIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGL PSPVDAAFEDAQGHIWFFQG corresponding to amino acids 165 - 359 of MM11_HUMAN, which also corresponds to amino acids 149 - 343 of HSSTROL3_P9, and a fourth amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TTGVSTPAPGV corresponding to amino acids 344 - 354 of HSSTROL3_P9, wherein said first amino acid sequence, second amino acid sequence, bridging amino acid, third amino acid sequence and fourth amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HSSTROL3_P9, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise KR, having a structure as follows: a sequence starting from any of amino acid numbers 96-x to 96; and ending at any of amino acid numbers 97+ ((n-2) - x), in which x varies from 0 to n-2.

3.An isolated polypeptide encoding for a tail of HSSTROL3_P9, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TTGVSTPAPGV in HSSTROL3_P9.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSSTROL3_P9 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1078, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1078 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
38	V -> A	Yes
198	A ->	No
307	Q -> H	Yes

Variant protein HSSTROL3_P9 is encoded by the following transcript(s): HSSTROL3_T12, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSSTROL3_T12 is shown in bold; this coding portion starts at position 24 and ends at position 1085. The transcript also has the following SNPs as listed in Table 1079 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSSTROL3_P9 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1079 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	T -> C	Yes
615	G ->	No
651	-> T	No
944	G -> C	Yes
1275	C ->	No
1327	C -> A	Yes
1353	A -> T	Yes
1570	-> G	No
1579	-> G	No
1593	C ->	No
1612	-> A	No
1638	T -> C	No
1696	A ->	No
1696	A -> C	No
1846	A ->	No
1846	A -> C	No
1876	T -> C	Yes
2023	-> T	No

As noted above, cluster HSSTROL3 features 16 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HSSTROL3_node_6 according to the present invention is supported by 14 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1080 below describes the starting and ending position of this segment on each transcript.

Table 1080 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	1	131
HSSTROL3_T8	1	131
HSSTROL3_T9	1	131
HSSTROL3_T10	1	131
HSSTROL3_T11	1	131
HSSTROL3_T12	1	131

Segment cluster HSSTROL3_node_10 according to the present invention is supported by 21 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1081 below describes the starting and ending position of this segment on each transcript.

Table 1081 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	132	313
HSSTROL3_T8	132	313
HSSTROL3_T9	132	313
HSSTROL3_T10	132	313
HSSTROL3_T11	132	313
HSSTROL3_T12	132	313

Segment cluster HSSTROL3_node_13 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1082 below describes the starting and ending position of this segment on each transcript.

Table 1082 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	362	505
HSSTROL3_T8	362	505
HSSTROL3_T9	362	505
HSSTROL3_T10	362	505
HSSTROL3_T11	362	505
HSSTROL3_T12	314	457

Segment cluster HSSTROL3_node_15 according to the present invention is supported by 47 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9,

HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1083 below describes the starting and ending position of this segment on each transcript.

Table 1083 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	506	639
HSSTROL3_T8	506	639
HSSTROL3_T9	506	639
HSSTROL3_T10	506	639
HSSTROL3_T11	506	639
HSSTROL3_T12	458	591

Segment cluster HSSTROL3_node_19 according to the present invention is supported by 63 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1084 below describes the starting and ending position of this segment on each transcript.

Table 1084 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	699	881
HSSTROL3_T8	699	881
HSSTROL3_T9	699	881
HSSTROL3_T10	699	881
HSSTROL3_T11	699	881
HSSTROL3_T12	651	833

Segment cluster HSSTROL3_node_21 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1085 below describes the starting and ending position of this segment on each transcript.

Table 1085 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	882	1098
HSSTROL3_T8	882	1098
HSSTROL3_T9	882	1098
HSSTROL3_T10	882	1098
HSSTROL3_T11	974	1190
HSSTROL3_T12	834	1050

Segment cluster HSSTROL3_node_24 according to the present invention is supported by 7 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T8 and HSSTROL3_T9. Table 1086 below describes the starting and ending position of this segment on each transcript.

Table 1086 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T8	1099	1236
HSSTROL3_T9	1099	1236

Segment cluster HSSTROL3_node_25 according to the present invention is supported by 13 libraries. The number of libraries was determined as previously described. This segment can

be found in the following transcript(s): HSSTROL3_T8. Table 1087 below describes the starting and ending position of this segment on each transcript.

Table 1087 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T8	1237	1536

Segment cluster HSSTROL3_node_26 according to the present invention is supported by 55 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9 and HSSTROL3_T11. Table 1088 below describes the starting and ending position of this segment on each transcript.

Table 1088 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	1099	1240
HSSTROL3_T8	1537	1678
HSSTROL3_T9	1237	1378
HSSTROL3_T11	1191	1332

Segment cluster HSSTROL3_node_28 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T9 and HSSTROL3_T10. Table 1089 below describes the starting and ending position of this segment on each transcript.

Table 1089 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HSSTROL3_T5	1357	2283
HSSTROL3_T9	1495	2421
HSSTROL3_T10	1215	2141

Segment cluster HSSTROL3_node_29 according to the present invention is supported by 109 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1090 below describes the starting and ending position of this segment on each transcript.

Table 1090 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	2284	3194
HSSTROL3_T8	1795	2705
HSSTROL3_T9	2422	3332
HSSTROL3_T10	2142	3052
HSSTROL3_T11	1449	2359
HSSTROL3_T12	1167	2077

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HSSTROL3_node_11 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10 and HSSTROL3_T11. Table 1091 below describes the starting and ending position of this segment on each transcript.

Table 1091 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	314	361
HSSTROL3_T8	314	361
HSSTROL3_T9	314	361
HSSTROL3_T10	314	361
HSSTROL3_T11	314	361

Segment cluster HSSTROL3_node_17 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1092 below describes the starting and ending position of this segment on each transcript.

Table 1092 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	640	680
HSSTROL3_T8	640	680
HSSTROL3_T9	640	680
HSSTROL3_T10	640	680
HSSTROL3_T11	640	680
HSSTROL3_T12	592	632

Segment cluster HSSTROL3_node_18 according to the present invention can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1093 below describes the starting and ending position of this segment on each transcript.

Table 1093 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T5	681	698
HSSTROL3_T8	681	698
HSSTROL3_T9	681	698
HSSTROL3_T10	681	698
HSSTROL3_T11	681	698
HSSTROL3_T12	633	650

Segment cluster HSSTROL3_node_20 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T11. Table 1094 below describes the starting and ending position of this segment on each transcript.

Table 1094 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSSTROL3_T11	882	973

Segment cluster HSSTROL3_node_27 according to the present invention is supported by 50 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSSTROL3_T5, HSSTROL3_T8, HSSTROL3_T9, HSSTROL3_T10, HSSTROL3_T11 and HSSTROL3_T12. Table 1095 below describes the starting and ending position of this segment on each transcript.

Table 1095 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

HSSTROL3_T5	1241	1356
HSSTROL3_T8	1679	1794
HSSTROL3_T9	1379	1494
HSSTROL3_T10	1099	1214
HSSTROL3_T11	1333	1448
HSSTROL3_T12	1051	1166

Variant protein alignment to the previously known protein:

Sequence name: MM11_HUMAN

Sequence documentation:

Alignment of: HSSTROL3_P4 x MM11_HUMAN ..

Alignment segment 1/1:

Quality: 4444.00

Escore: 0

Matching length: 445 Total length: 445

Matching Percent Similarity: 99.78 Matching Percent Identity: 99.78

Total Percent Similarity: 99.78 Total Percent Identity: 99.78

Gaps: 0

Alignment:

```
                 .         .         .         .         .
       1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
                 .         .         .         .         .
      51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
                 .         .         .         .         .
     101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
                 .         .         .         .         .
     151 GRADIMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
         |||||||||||| ||||||||||||||||||||||||||||||||||||||
     151 GRADIMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
                 .         .         .         .         .
     201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
                 .         .         .         .         .
     251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
                 .         .         .         .         .
     301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350
                 .         .         .         .         .
     351 QGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRFPVHAALVWGPEKNKI 400
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||
351 QGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRFPVHAALVWGPEKNKI 400
                          .         .         .         .
401 YFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADG       445
                ||||||||||||||||||||||||||||||||||||||||||
401 YFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADG       445
```

Sequence name: MM11_HUMAN

Sequence documentation:

Alignment of: HSSTROL3_P5 x MM11_HUMAN ..

Alignment segment 1/1:

Quality: 3566.00

Escore: 0

Matching length: 358 Total length: 358

Matching Percent Similarity: 99.72 Matching Percent Identity: 99.72

Total Percent Similarity: 99.72 Total Percent Identity: 99.72

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
                .         .         .         .         .
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
                .         .         .         .         .
101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
                .         .         .         .         .
151 GRADIMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
    |||||||||||| ||||||||||||||||||||||||||||||||||||
151 GRADIMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
                .         .         .         .         .
201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
                .         .         .         .         .
251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
                .         .         .         .         .
301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350
                                                        .
351 QGHIWFFQ                                            358
    ||||||||
```

351 QGHIWFFQ 358

Sequence name: MM11_HUMAN

Sequence documentation:

Alignment of: HSSTROL3_P7 x MM11_HUMAN ..

Alignment segment 1/1:

Quality: 3575.00

Escore: 0

Matching length: 359 Total length: 359

Matching Percent Similarity: 99.72 Matching Percent Identity: 99.72

Total Percent Similarity: 99.72 Total Percent Identity: 99.72

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
                .         .         .         .         .
```

```
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
            .         .         .         .         .
101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
            .         .         .         .         .
151 GRADIMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
    |||||||||||| ||||||||||||||||||||||||||||||||||||
151 GRADIMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
            .         .         .         .         .
201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
            .         .         .         .         .
251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
            .         .         .         .         .
301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350

351 QGHIWFFQG                                          359
    |||||||||
351 QGHIWFFQG                                          359
```

Sequence name: MM11_HUMAN

Sequence documentation:

Alignment of: HSSTROL3_P8 x MM11_HUMAN ..

Alignment segment 1/1:

Quality: 2838.00

Escore: 0

Matching length: 286 Total length: 286

Matching Percent Similarity: 99.65 Matching Percent Identity: 99.65

Total Percent Similarity: 99.65 Total Percent Identity: 99.65

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
                .         .         .         .         .
       51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
                .         .         .         .         .
      101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
          101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
                        .         .         .         .         .
          151 GRADIMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
                ||||||||||||| ||||||||||||||||||||||||||||||||||||
          151 GRADIMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
                        .         .         .         .         .
          201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
          201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
                        .         .         .
          251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLE               286
                ||||||||||||||||||||||||||||||||||
          251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLE               286
```

Sequence name: MM11_HUMAN

Sequence documentation:

Alignment of: HSSTROL3_P9 x MM11_HUMAN ..

Alignment segment 1/1:

Quality: 3316.00

Escore: 0

Matching length: 343 Total length: 359

Matching Percent Similarity: 99.71 Matching Percent Identity: 99.71

Total Percent Similarity: 95.26 Total Percent Identity: 95.26

Gaps: 1

Alignment:

```
                .         .         .         .         .
  1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQP 50
                .         .         .         .         .
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQK.... 96
    ||||||||||||||||||||||||||||||||||||||||||||||
 51 WHAALPSSPAPAPATQEAPRPASSLRPPRCGVPDPSDGLSARNRQKRFVL 100
                .         .         .         .         .
 97 ............RILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 134
                  ||||||||||||||||||||||||||||||||||||
101 SGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHE 150
                .         .         .         .         .
135 GRADIMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 184
    |||||||||||| ||||||||||||||||||||||||||||||||||||
151 GRADIMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWT 200
                .         .         .         .         .
185 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 IGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSPDDC 250
                .         .         .         .         .
235 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 284
```

1130

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
          251 RGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDA 300
                          .         .         .         .         .
          285 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 334
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
          301 VSTIRGELFFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDA 350

          335 QGHIWFFQG                                           343
                |||||||||
          351 QGHIWFFQG                                           359
```

Expression of Stromelysin-3 precursor HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 seg24 in normal and cancerous Lung tissues

Expression of Stromelysin-3 precursor (EC 3.4.24.-) (Matrix metalloproteinase-11) (MMP-11) (ST3) (SL-3) transcripts detectable by or according to seg24, HSSTROL3 seg24 amplicon (SEQ ID NO: 1675) and HSSTROL3 seg24F (SEQ ID NO: 1673) and HSSTROL3 seg24R (SEQ ID NO: 1674) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin(GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2 "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 39 is a histogram showing over expression of the above-indicated Stromelysin-3 precursor transcripts in cancerous lung samples relative to the normal samples. Values

represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.)

As is evident from Figure 39, the expression of Stromelysin-3 precursor transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 13 out of 15 adenocarcinoma samples, 8 out of 16 squamous cell carcinoma samples, 3 out of 4 large cell carcinoma samples and in 7 out of 8 small cell carcinoma samples.

Threshold of 5 fold overexpression was found to differentiate between cancer and normal samples with P value of 4.04E-04 in adenocarcinoma, 9.89E-02 in squamous cell carcinoma, 6.04E-02 in Large cell carcinoma, 3.14E-03 in small cell carcinoma as checked by exact fisher test. The above values demonstrate statistical significance of the results.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: HSSTROL3 seg24F forward primer; and HSSTROL3 seg24R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: HSSTROL3 seg24.

Forward Primer (SEQ ID NO: 1673): ATTTCCATCCTCAACTGGCAGA

Reverse Primer (SEQ ID NO: 1674): TGCCCTGGAACCCACG

Amplicon (SEQ ID NO: 1675):

ATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCCAGGGAC TTCACAAATGAAGGCACAGCATGGGAAACCTGCGTGGGTTCCAGGGCA

Expression of Stromelysin-3 precursor HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 seg24 in different normal tissues

Expression of Stromelysin-3 precursor transcripts detectable by or according to HSSTROL3 seg24 amplicon (SEQ ID NO: 1675) and HSSTROL3 seg24F (SEQ ID NO: 1673)

and HSSTROL3 seg24R (SEQ ID NO: 1674) was measured by real time PCR. In parallel the expression of four housekeeping genes Ubiquitin(GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the lung samples (Sample Nos. 15-17, Table 2 "Tissue samples in normal panel", above), to obtain a value of relative expression of each sample relative to median of the lung samples.

Forward Primer (SEQ ID NO: 1673): ATTTCCATCCTCAACTGGCAGA

Reverse Primer (SEQ ID NO: 1674): TGCCCTGGAACCCACG

Amplicon (SEQ ID NO: 1675):

ATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCCAGGGAC
TTCACAAATGAAGGCACAGCATGGGAAACCTGCGTGGGTTCCAGGGCA

The results are demonstrated in Figure 40, showing the expression of Stromelysin-3 HSSTROL3 transcripts, which are detectable by amplicon as depicted in sequence name HSSTROL3 seg24, in different normal tissues.

Expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) *HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 seg20-21 in normal and cancerous lung tissues*

Expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts detectable by or according to seg20-21, HSSTROL3 seg20-21 amplicon (SEQ ID NO: 1678) and primers HSSTROL3 seg20-21F (SEQ ID NO: 1676) and HSSTROL3 seg20-21R (SEQ ID NO: 1677) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-

amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 71 is a histogram showing over expression of the above-indicated Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 71, the expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2,). Notably an over-expression of at least 6 fold was found in 11 out of 15 adenocarcinoma samples, 6 out of 16 squamous cell carcinoma samples, 1 out of 4 large cell carcinoma samples and in 6 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: HSSTROL3 seg20-21F forward primer; and HSSTROL3 seg20-21R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: HSSTROL3 seg20-21.

Primers:

Forward primer HSSTROL3 seg20-21F (SEQ ID NO: 1676): TCTGCTGGCCACTGTGACTG
Reverse primer HSSTROL3 seg20-21R (SEQ ID NO: 1677):
GAAGAAAAAGAGCTCGCCTCG
Amplicon HSSTROL3 seg20-21 (SEQ ID NO: 1678):
TCTGCTGGCCACTGTGACTGCAGCATATGCCCTCAGCATGTGTCCCTCTCTCCCACC
CCAGCCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGGTCTCCACCA
TCCGAGGCGAGCTCTTTTTCTTC

Expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) *HSSTROL3 transcripts which are detectable by amplicon as depicted in sequence name HSSTROL3 junc21-27 in normal and cancerous lung tissues*

Expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts detectable by or according to junc21-27, HSSTROL3 junc21-27 amplicon (SEQ ID NO: 1681) and primers HSSTROL3 junc21-27F (SEQ ID NO: 1679) and HSSTROL3 junc21-27R (SEQ ID NO: 1680) was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-

amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 72 is a histogram showing over expression of the above-indicated Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 72, the expression of Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11) transcripts detectable by the above amplicon(s) in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2,). Notably an over-expression of at least 10 fold was found in 15 out of 15 adenocarcinoma samples, 13 out of 16 squamous cell carcinoma samples, 3 out of 4 large cell carcinoma samples and in 5 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: HSSTROL3 junc21-27F forward primer; and HSSTROL3 junc21-27R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: HSSTROL3 junc21-27.

Primers:

Forward primer HSSTROL3 junc21-27F (SEQ ID NO: 1679):
ACATTTGGTTCTTCCAAGGGACTAC

Reverse primer HSSTROL3 junc21-27R (SEQ ID NO: 1680):
TCGATCTCAGAGGGCACCC

Amplicon HSSTROL3 junc21-27 (SEQ ID NO: 1681):
ACATTTGGTTCTTCCAAGGGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGA
CAGTCCCGTGCCCCGCAGGGCCACTGACTGGAGAGGGGTGCCCTCTGAGATCGA

1137

DESCRIPTION FOR CLUSTER HUMTREFAC

Cluster HUMTREFAC features 2 transcript(s) and 7 segment(s) of interest, the names for which are given in Tables 1096 and 1097, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1098.

Table 1096 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMTREFAC_PEA_2_T4	131
HUMTREFAC_PEA_2_T5	132

Table 1097 - Segments of interest

Segment Name	Sequence ID No.
HUMTREFAC_PEA_2_node_0	903
HUMTREFAC_PEA_2_node_9	904
HUMTREFAC_PEA_2_node_2	905
HUMTREFAC_PEA_2_node_3	906
HUMTREFAC_PEA_2_node_4	907
HUMTREFAC_PEA_2_node_5	908
HUMTREFAC_PEA_2_node_8	909

Table 1098 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HUMTREFAC_PEA_2_P7	1399	HUMTREFAC_PEA_2_T5
HUMTREFAC_PEA_2_P8	1400	HUMTREFAC_PEA_2_T4

These sequences are variants of the known protein Trefoil factor 3 precursor (SwissProt accession identifier TFF3_HUMAN; known also according to the synonyms Intestinal trefoil factor; hP1.B), SEQ ID NO: 1456, referred to herein as the previously known protein.

Protein Trefoil factor 3 precursor is known or believed to have the following function(s): May have a role in promoting cell migration (motogen). The sequence for protein Trefoil factor

3 precursor is given at the end of the application, as "Trefoil factor 3 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1099.

Table 1099 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
74 - 76	QEA -> TRKT

Protein Trefoil factor 3 precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: defense response; digestion, which are annotation(s) related to Biological Process; and extracellular, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HUMTREFAC can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 41 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 41 and Table 1100. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: a mixture of malignant tumors from different tissues, breast malignant tumors, pancreas carcinoma and prostate cancer.

Table 1100 - Normal tissue distribution

Name of Tissue	Number

adrenal	40
colon	797
epithelial	95
general	39
liver	0
lung	57
lymph nodes	3
breast	0
muscle	3
pancreas	2
prostate	16
stomach	0
Thyroid	257
uterus	54

Table 1101 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	6.4e-01	6.9e-01	7.1e-01	1.1	7.8e-01	0.9
colon	4.6e-01	5.7e-01	9.7e-01	0.5	1	0.4
epithelial	2.4e-02	3.4e-01	9.5e-10	2.0	5.3e-02	1.1
general	2.5e-04	3.9e-02	1.4e-28	3.6	1.9e-10	1.9
liver	1	6.8e-01	1	1.0	6.9e-01	1.4
lung	4.8e-01	7.6e-01	2.2e-03	1.0	1.6e-01	0.5
lymph nodes	5.1e-01	8.0e-01	2.3e-02	5.0	1.9e-01	2.1
breast	7.6e-02	1.2e-01	3.1e-06	12.0	1.1e-03	6.5
muscle	9.2e-01	4.8e-01	1	0.8	3.9e-01	2.1
pancreas	1.2e-01	2.4e-01	5.7e-03	6.5	2.1e-02	4.6
prostate	1.5e-01	2.7e-01	9.9e-10	8.1	3.1e-07	5.7
stomach	3.0e-01	1.3e-01	5.0e-01	2.0	6.7e-02	2.8

Thyroid	6.4e-01	6.4e-01	9.6e-01	0.5	9.6e-01	0.5
uterus	4.1e-01	7.3e-01	7.5e-02	1.3	4.0e-01	0.8

As noted above, cluster HUMTREFAC features 2 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Trefoil factor 3 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HUMTREFAC_PEA_2_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMTREFAC_PEA_2_T5. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMTREFAC_PEA_2_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1102, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMTREFAC_PEA_2_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1102 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> S	No
5	A -> T	No
14	A -> V	Yes
43	L -> M	No

60	P -> S	Yes
123	S -> *	Yes

Variant protein HUMTREFAC_PEA_2_P7 is encoded by the following transcript(s): HUMTREFAC_PEA_2_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMTREFAC_PEA_2_T5 is shown in bold; this coding portion starts at position 278 and ends at position 688. The transcript also has the following SNPs as listed in Table 1103 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMTREFAC_PEA_2_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1103 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
233	A -> G	Yes
290	G -> A	No
290	G -> T	No
318	C -> T	Yes
404	C -> A	No
404	C -> T	No
455	C -> T	Yes
645	C -> A	Yes
685	C -> T	No

Variant protein HUMTREFAC_PEA_2_P8 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMTREFAC_PEA_2_T4. An alignment is given to the known protein (Trefoil factor 3 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the

relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMTREFAC_PEA_2_P8 and TFF3_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMTREFAC_PEA_2_P8, comprising a first amino acid sequence being at least 90 % homologous to MAARALCMLGLVLALLSSSSAEEYVGL corresponding to amino acids 1 - 27 of TFF3_HUMAN, which also corresponds to amino acids 1 - 27 of HUMTREFAC_PEA_2_P8, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence WKVHLPKGEGFSSG corresponding to amino acids 28 - 41 of HUMTREFAC_PEA_2_P8, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMTREFAC_PEA_2_P8, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence WKVHLPKGEGFSSG in HUMTREFAC_PEA_2_P8.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMTREFAC_PEA_2_P8 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1104, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMTREFAC_PEA_2_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1104 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
5	A -> S	No
5	A -> T	No
14	A -> V	Yes

Variant protein HUMTREFAC_PEA_2_P8 is encoded by the following transcript(s): HUMTREFAC_PEA_2_T4, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMTREFAC_PEA_2_T4 is shown in bold; this coding portion starts at position 278 and ends at position 400. The transcript also has the following SNPs as listed in Table 1105 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMTREFAC_PEA_2_P8 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1105 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
233	A -> G	Yes
290	G -> A	No
290	G -> T	No
318	C -> T	Yes
515	C -> A	No
515	C -> T	No
566	C -> T	Yes
756	C -> A	Yes
796	C -> T	No
1265	A -> C	No
1266	A -> T	No

As noted above, cluster HUMTREFAC features 7 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s)

are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMTREFAC_PEA_2_node_0 according to the present invention is supported by 188 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1106 below describes the starting and ending position of this segment on each transcript.

Table 1106 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	1	359
HUMTREFAC_PEA_2_T5	1	359

Segment cluster HUMTREFAC_PEA_2_node_9 according to the present invention is supported by 150 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1107 below describes the starting and ending position of this segment on each transcript.

Table 1107 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	681	1266
HUMTREFAC_PEA_2_T5	570	747

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMTREFAC_PEA_2_node_2 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4. Table 1108 below describes the starting and ending position of this segment on each transcript.

Table 1108 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	360	470

Segment cluster HUMTREFAC_PEA_2_node_3 according to the present invention is supported by 10 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1109 below describes the starting and ending position of this segment on each transcript.

Table 1109 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	471	514
HUMTREFAC_PEA_2_T5	360	403

Segment cluster HUMTREFAC_PEA_2_node_4 according to the present invention is supported by 197 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1110 below describes the starting and ending position of this segment on each transcript.

Table 1110 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	515	611
HUMTREFAC_PEA_2_T5	404	500

Segment cluster HUMTREFAC_PEA_2_node_5 according to the present invention is supported by 187 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1111 below describes the starting and ending position of this segment on each transcript.

Table 1111 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	612	661
HUMTREFAC_PEA_2_T5	501	550

Segment cluster HUMTREFAC_PEA_2_node_8 according to the present invention can be found in the following transcript(s): HUMTREFAC_PEA_2_T4 and HUMTREFAC_PEA_2_T5. Table 1112 below describes the starting and ending position of this segment on each transcript.

Table 1112 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMTREFAC_PEA_2_T4	662	680
HUMTREFAC_PEA_2_T5	551	569

Variant protein alignment to the previously known protein:

Sequence name: TFF3_HUMAN

Sequence documentation:

Alignment of: HUMTREFAC_PEA_2_P8 x TFF3_HUMAN ..

Alignment segment 1/1:

Quality: 246.00

Escore: 0

Matching length: 27 Total length: 27

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                    .         .
        1 MAARALCMLGLVLALLSSSSAEEYVGL                          27
          |||||||||||||||||||||||||||
        1 MAARALCMLGLVLALLSSSSAEEYVGL                          27
```

DESCRIPTION FOR CLUSTER HSS100PCB

Cluster HSS100PCB features 1 transcript(s) and 3 segment(s) of interest, the names for which are given in Tables 1113 and 1114, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1115.

Table 1113 - Transcripts of interest

Transcript Name	Sequence ID No.
HSS100PCB_T1	133

Table 1114 - Segments of interest

Segment Name	Sequence ID No.
HSS100PCB_node_3	910
HSS100PCB_node_4	911
HSS100PCB_node_5	912

Table 1115 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HSS100PCB_P3	1401	HSS100PCB_T1

These sequences are variants of the known protein S-100P protein (SwissProt accession identifier S10P_HUMAN), SEQ ID NO:1457, referred to herein as the previously known protein, which binds two calcium ions.

The sequence for protein S-100P protein is given at the end of the application, as "S-100P protein amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1116.

Table 1116 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
32	E -> T
44	F -> E

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: calcium binding; protein binding, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HSS100PCB can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 42 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 42 and Table 1117. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: a mixture of malignant tumors from different tissues.

Table 1117 - Normal tissue distribution

Name of Tissue	Number
bladder	41
colon	37
epithelial	38
general	22
kidney	0
liver	0
lung	18
breast	0
bone marrow	0
ovary	0
pancreas	0

prostate	46
stomach	553
uterus	13

Table 1118 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bladder	3.3e-01	2.9e-01	2.9e-02	2.8	3.5e-02	2.8
colon	3.0e-01	1.9e-01	5.2e-01	1.2	2.4e-01	1.7
epithelial	4.7e-02	1.6e-02	2.0e-01	1.2	6.1e-02	1.3
general	1.1e-03	6.8e-05	1.4e-02	1.5	4.9e-04	1.7
kidney	6.5e-01	7.2e-01	5.8e-01	1.7	7.0e-01	1.4
liver	9.1e-01	4.9e-01	1	1.0	7.7e-02	2.1
lung	6.8e-01	7.3e-01	2.2e-02	2.9	1.3e-01	1.7
breast	2.8e-01	3.2e-01	4.7e-01	2.0	6.8e-01	1.5
bone marrow	1	6.7e-01	1	1.0	2.8e-01	2.8
ovary	2.6e-01	3.0e-01	4.7e-01	2.0	5.9e-01	1.7
pancreas	3.3e-01	4.4e-01	7.6e-02	3.7	1.5e-01	2.8
prostate	9.1e-01	9.3e-01	5.8e-01	0.6	7.6e-01	0.5
stomach	3.7e-01	3.2e-01	1	0.1	1	0.3
uterus	9.4e-01	7.0e-01	1	0.6	4.1e-01	1.1

As noted above, cluster HSS100PCB features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein S-100P protein. A description of each variant protein according to the present invention is now provided.

Variant protein HSS100PCB_P3 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSS100PCB_T1. The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide

prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HSS100PCB_P3 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1119, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSS100PCB_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1119 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
1	M -> R	Yes
11	M -> L	Yes
20	L -> F	Yes

Variant protein HSS100PCB_P3 is encoded by the following transcript(s): HSS100PCB_T1, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSS100PCB_T1 is shown in bold; this coding portion starts at position 1057 and ends at position 1533. The transcript also has the following SNPs as listed in Table 1120 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSS100PCB_P3 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1120 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
52	C -> T	Yes
107	A -> C	Yes
458	C -> T	Yes

468	A -> G	Yes
648	C -> T	Yes
846	C -> G	Yes
882	G -> A	Yes
960	C -> T	No
965	C -> T	Yes
1058	T -> G	Yes
1087	A -> C	Yes
1114	C -> T	Yes
1968	G -> A	Yes
1971	C -> T	Yes
2010	C -> A	Yes
2099	G ->	No

As noted above, cluster HSS100PCB features 3 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HSS100PCB_node_3 according to the present invention is supported by 16 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSS100PCB_T1. Table 1121 below describes the starting and ending position of this segment on each transcript.

Table 1121 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSS100PCB_T1	1	1133

Segment cluster HSS100PCB_node_4 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSS100PCB_T1. Table 1122 below describes the starting and ending position of this segment on each transcript.

Table 1123 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSS100PCB_T1	1134	1923

Segment cluster HSS100PCB_node_5 according to the present invention is supported by 141 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSS100PCB_T1. Table 1124 below describes the starting and ending position of this segment on each transcript.

Table 1124 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSS100PCB_T1	1924	2201

1154

DESCRIPTION FOR CLUSTER HSU33147

Cluster HSU33147 features 2 transcript(s) and 5 segment(s) of interest, the names for which are given in Tables 1125 and 1126, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1127.

Table 1125 - Transcripts of interest

Transcript Name	Sequence ID No.
HSU33147_PEA_1_T1	1464
HSU33147_PEA_1_T2	1465

Table 1126 - Segments of interest

Segment Name	Sequence ID No.
HSU33147_PEA_1_node_0	1276
HSU33147_PEA_1_node_2	1277
HSU33147_PEA_1_node_4	1278
HSU33147_PEA_1_node_7	1279
HSU33147_PEA_1_node_3	1280

Table 1127 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HSU33147_PEA_1_P5	1415	HSU33147_PEA_1_T1; HSU33147_PEA_1_T2

These sequences are variants of the known protein Mammaglobin A precursor (SwissProt accession identifier MGBA_HUMAN; known also according to the synonyms Mammaglobin 1; Secretoglobin family 2A member 2), SEQ ID NO: 1416, referred to herein as the previously known protein.

The sequence for protein Mammaglobin A precursor is given at the end of the application, as "Mammaglobin A precursor amino acid sequence".

It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Immunostimulant. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Anticancer.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: steroid binding, which are annotation(s) related to Molecular Function.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HSU33147 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the left hand column of the table and the numbers on the y-axis of figure 43 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 43 and Table 1128. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: a mixture of malignant tumors from different tissues.

Table 1128 - Normal tissue distribution

Name of Tissue	Number
epithelial	6
general	2

lung	0
breast	131

Table 1129 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
epithelial	4.1e-02	6.4e-02	1.5e-12	2.6	2.2e-06	1.5
general	1.6e-02	1.1e-02	1.2e-22	4.4	7.2e-13	2.4
lung	1	6.3e-01	1	1.0	6.2e-01	1.6
breast	8.6e-02	1.1e-01	3.4e-07	1.7	2.6e-03	1.0

As noted above, cluster HSU33147 features 2 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Mammaglobin A precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HSU33147_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HSU33147_PEA_1_T1. An alignment is given to the known protein (Mammaglobin A precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HSU33147_PEA_1_P5 and MGBA_HUMAN:

1.An isolated chimeric polypeptide encoding for HSU33147_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAI DELKECFLNQTDETLSNVE corresponding to amino acids 1 - 78 of MGBA_HUMAN, which also corresponds to amino acids 1 - 78 of HSU33147_PEA_1_P5, and a second amino acid sequence being at least 90 % homologous to QLIYDSSLCDLF corresponding to amino acids 82 - 93 of MGBA_HUMAN, which also corresponds to amino acids 79 - 90 of

HSU33147_PEA_1_P5, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of HSU33147_PEA_1_P5, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise EQ, having a structure as follows: a sequence starting from any of amino acid numbers 78-x to 78; and ending at any of amino acid numbers 79+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

The glycosylation sites of variant protein HSU33147_PEA_1_P5, as compared to the known protein Mammaglobin A precursor, are described in Table 1130 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1130 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
68	yes	68
53	yes	53

Variant protein HSU33147_PEA_1_P5 is encoded by the following transcript(s): HSU33147_PEA_1_T1, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HSU33147_PEA_1_T1 is shown in bold; this coding portion starts at position 72 and ends at position 341. The transcript also has the following SNPs as listed in Table 1131 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HSU33147_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1131 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
84	A -> C	No
124	C ->	No
396	A -> G	No

As noted above, cluster HSU33147 features 5 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HSU33147_PEA_1_node_0 according to the present invention is supported by 38 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSU33147_PEA_1_T1 and HSU33147_PEA_1_T2. Table 1132 below describes the starting and ending position of this segment on each transcript.

Table 1132 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSU33147_PEA_1_T1	1	126

HSU33147_PEA_1_T2	1	126

Segment cluster HSU33147_PEA_1_node_2 according to the present invention is supported by 44 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSU33147_PEA_1_T1 and HSU33147_PEA_1_T2. Table 1133 below describes the starting and ending position of this segment on each transcript.

Table 1133 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSU33147_PEA_1_T1	127	305
HSU33147_PEA_1_T2	127	305

Segment cluster HSU33147_PEA_1_node_4 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSU33147_PEA_1_T2. Table 1134 below describes the starting and ending position of this segment on each transcript.

Table 1134 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSU33147_PEA_1_T2	315	907

Segment cluster HSU33147_PEA_1_node_7 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HSU33147_PEA_1_T1. Table 1135 below describes the starting and ending position of this segment on each transcript.

Table 1135 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSU33147_PEA_1_T1	306	516

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HSU33147_PEA_1_node_3 according to the present invention can be found in the following transcript(s): HSU33147_PEA_1_T2. Table 1136 below describes the starting and ending position of this segment on each transcript.

Table 1136 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HSU33147_PEA_1_T2	306	314

Variant protein alignment to the previously known protein:

Sequence name: MGBA_HUMAN

Sequence documentation:

Alignment of: HSU33147_PEA_1_P5 x MGBA_HUMAN ..

Alignment segment 1/1:

Quality: 776.00

Escore: 0

Matching length: 90 Total length: 93

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 96.77 Total Percent Identity: 96.77

Gaps: 1

Alignment:

```
                .         .         .         .         .
        1 MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFI 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFI 50
                .         .         .         .
       51 DDNATTNAIDELKECFLNQTDETLSNVE...QLIYDSSLCDLF        90
          ||||||||||||||||||||||||||||   ||||||||||||
       51 DDNATTNAIDELKECFLNQTDETLSNVEVFMQLIYDSSLCDLF        93
```

DESCRIPTION FOR CLUSTER R20779

Cluster R20779 features 1 transcript(s) and 24 segment(s) of interest, the names for which are given in Tables 1137 and 1138, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1139.

Table 1137 - Transcripts of interest

Transcript Name	Sequence ID No.
R20779_T7	134

Table 1138 - Segments of interest

Segment Name	Sequence ID No.
R20779_node_0	913
R20779_node_2	914
R20779_node_7	915
R20779_node_9	916
R20779_node_18	917
R20779_node_21	918
R20779_node_24	919
R20779_node_27	920
R20779_node_28	921
R20779_node_30	922
R20779_node_31	923
R20779_node_32	924
R20779_node_1	925
R20779_node_3	926
R20779_node_10	927
R20779_node_11	928
R20779_node_14	929
R20779_node_17	930
R20779_node_19	931
R20779_node_20	932
R20779_node_22	933
R20779_node_23	934
R20779_node_25	935
R20779_node_29	936

Table 1139 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)

R20779_P2	1402	R20779_T7

These sequences are variants of the known protein Stanniocalcin 2 precursor (SwissProt accession identifier STC2_HUMAN; known also according to the synonyms STC-2; Stanniocalcin-related protein; STCRP; STC-related protein), SEQ ID NO:1458, referred to herein as the previously known protein.

Protein Stanniocalcin 2 precursor is known or believed to have the following function(s): Has an anti-hypocalcemic action on calcium and phosphate homeostasis. The sequence for protein Stanniocalcin 2 precursor is given at the end of the application, as "Stanniocalcin 2 precursor amino acid sequence". Protein Stanniocalcin 2 precursor localization is believed to be Secreted (Potential).

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: cell surface receptor linked signal transduction; cell-cell signaling; nutritional response pathway, which are annotation(s) related to Biological Process; hormone, which are annotation(s) related to Molecular Function; and extracellular, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster R20779 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 44 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 44 and Table 1140. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues and lung malignant tumors.

Table 1140 - Normal tissue distribution

Name of Tissue	Number
bone	825
brain	0
colon	0
epithelial	32
general	38
kidney	22
liver	9
lung	11
lymph nodes	0
breast	215
muscle	35
ovary	36
pancreas	4
prostate	80
skin	99
stomach	0
uterus	4

Table 1141 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
bone	5.9e-01	7.4e-01	1	0.2	1	0.1
brain	2.5e-02	1.6e-02	2.2e-01	6.0	3.5e-02	8.0
colon	1.7e-01	1.7e-01	1	1.3	7.7e-01	1.5
epithelial	1.7e-01	1.5e-03	5.9e-01	1.0	2.0e-04	2.0
general	2.4e-02	6.2e-07	7.6e-01	0.8	4.6e-05	1.6
kidney	4.3e-01	2.7e-01	6.2e-01	1.3	1.5e-01	2.0

liver	8.3e-01	7.6e-01	1	0.8	3.3e-01	1.6
lung	1.2e-01	1.4e-03	1.9e-01	2.9	1.6e-05	7.7
lymph nodes	1	3.1e-01	1	1.0	1	1.4
breast	6.8e-01	6.8e-01	6.9e-01	0.8	3.6e-01	0.8
muscle	9.2e-01	4.8e-01	1	0.3	1.4e-03	1.4
ovary	8.4e-01	7.1e-01	9.0e-01	0.7	8.6e-01	0.8
pancreas	9.3e-01	6.8e-01	1	0.7	1.5e-01	2.0
prostate	9.1e-01	5.0e-01	9.8e-01	0.4	5.7e-01	0.7
skin	6.3e-01	7.5e-01	7.1e-01	0.8	9.5e-01	0.3
stomach	1	4.5e-01	1	1.0	5.1e-01	1.8
uterus	7.1e-01	2.6e-01	4.4e-01	1.7	4.1e-01	1.8

As noted above, cluster R20779 features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Stanniocalcin 2 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein R20779_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R20779_T7. An alignment is given to the known protein (Stanniocalcin 2 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R20779_P2 and STC2_HUMAN:

1.An isolated chimeric polypeptide encoding for R20779_P2, comprising a first amino acid sequence being at least 90 % homologous to

MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNTAEIQHCLV NAGDVGCGVFECFENNSCEIRGLHGICMTFLHNAGKFDAQGKSFIKDALKCKAHALRH RFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQENTRVIVEMIHFKDLLLHE

corresponding to amino acids 1 - 169 of STC2_HUMAN, which also corresponds to amino acids 1 - 169 of R20779_P2, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least

95% homologous to a polypeptide having the sequence CYKIEITMPKRRKVKLRD corresponding to amino acids 170 - 187 of R20779_P2, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R20779_P2, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence CYKIEITMPKRRKVKLRD in R20779_P2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R20779_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1142, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R20779_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1142 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
16	L ->	No
98	Q ->	No
171	Y -> C	Yes
177	M -> V	Yes

The glycosylation sites of variant protein R20779_P2, as compared to the known protein Stanniocalcin 2 precursor, are described in Table 1143 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the

glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1143 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
73	yes	73

Variant protein R20779_P2 is encoded by the following transcript(s): R20779_T7, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R20779_T7 is shown in bold; this coding portion starts at position 1397 and ends at position 1957. The transcript also has the following SNPs as listed in Table 1144 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R20779_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1144 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
1442	T ->	No
1690	G ->	No
1732	C -> T	Yes
1867	G -> T	Yes
1908	A -> G	Yes
1925	A -> G	Yes
1968	G -> A	Yes
2087	C -> T	No
2138	C -> T	Yes
2270	C ->	No
2443	A ->	No

2478	G ->	No
2479	C -> A	No
2616	C -> A	No
2941	C ->	No
3196	-> A	No
3479	T -> G	Yes
4290	C -> T	Yes
4358	G -> A	Yes
5363	G -> A	No

As noted above, cluster R20779 features 24 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R20779_node_0 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1145 below describes the starting and ending position of this segment on each transcript.

Table 1145 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1	1298

Segment cluster R20779_node_2 according to the present invention is supported by 55 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1146 below describes the starting and ending position of this segment on each transcript.

Table 1146 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1337	1506

Segment cluster R20779_node_7 according to the present invention is supported by 63 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1147 below describes the starting and ending position of this segment on each transcript.

Table 1147 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1548	1690

Segment cluster R20779_node_9 according to the present invention is supported by 66 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1148 below describes the starting and ending position of this segment on each transcript.

Table 1148 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1691	1838

Segment cluster R20779_node_18 according to the present invention is supported by 61 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1149 below describes the starting and ending position of this segment on each transcript.

Table 1149 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2009	2176

Segment cluster R20779_node_21 according to the present invention is supported by 106 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1150 below describes the starting and ending position of this segment on each transcript.

Table 1150 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2219	2796

Segment cluster R20779_node_24 according to the present invention is supported by 100 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1151 below describes the starting and ending position of this segment on each transcript.

Table 1151 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2977	3667

Segment cluster R20779_node_27 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be

found in the following transcript(s): R20779_T7. Table 1152 below describes the starting and ending position of this segment on each transcript.

Table 1152 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	3673	3803

Segment cluster R20779_node_28 according to the present invention is supported by 31 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1153 below describes the starting and ending position of this segment on each transcript.

Table 1153- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	3804	4050

Segment cluster R20779_node_30 according to the present invention is supported by 34 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1154 below describes the starting and ending position of this segment on each transcript.

Table 1154 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	4068	4193

Segment cluster R20779_node_31 according to the present invention is supported by 46 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1155 below describes the starting and ending position of this segment on each transcript.

Table 1155 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	4194	4424

Segment cluster R20779_node_32 according to the present invention is supported by 88 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1156 below describes the starting and ending position of this segment on each transcript.

Table 1156 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	4425	5503

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R20779_node_1 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1157 below describes the starting and ending position of this segment on each transcript.

Table 1157 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

R20779_T7	1299	1336

Segment cluster R20779_node_3 according to the present invention is supported by 52 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1158 below describes the starting and ending position of this segment on each transcript.

Table 1158 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1507	1547

Segment cluster R20779_node_10 according to the present invention can be found in the following transcript(s): R20779_T7. Table 1159 below describes the starting and ending position of this segment on each transcript.

Table 1159 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1839	1849

Segment cluster R20779_node_11 according to the present invention is supported by 58 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1160 below describes the starting and ending position of this segment on each transcript.

Table 1160 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

R20779_T7	1850	1902

Segment cluster R20779_node_14 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1161 below describes the starting and ending position of this segment on each transcript.

Table 1161 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1903	1975

Segment cluster R20779_node_17 according to the present invention is supported by 54 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1162 below describes the starting and ending position of this segment on each transcript.

Table 1162 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	1976	2008

Segment cluster R20779_node_19 according to the present invention can be found in the following transcript(s): R20779_T7. Table 1163 below describes the starting and ending position of this segment on each transcript.

Table 1163 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

R20779_T7	2177	2188

Segment cluster R20779_node_20 according to the present invention is supported by 53 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1164 below describes the starting and ending position of this segment on each transcript.

Table 1164 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2189	2218

Segment cluster R20779_node_22 according to the present invention is supported by 76 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1165 below describes the starting and ending position of this segment on each transcript.

Table 1165 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2797	2899

Segment cluster R20779_node_23 according to the present invention is supported by 81 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R20779_T7. Table 1166 below describes the starting and ending position of this segment on each transcript.

Table 1166 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	2900	2976

Segment cluster R20779_node_25 according to the present invention can be found in the following transcript(s): R20779_T7. Table 1167 below describes the starting and ending position of this segment on each transcript.

Table 1167 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	3668	3672

Segment cluster R20779_node_29 according to the present invention can be found in the following transcript(s): R20779_T7. Table 1168 below describes the starting and ending position of this segment on each transcript.

Table 1168 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R20779_T7	4051	4067

Variant protein alignment to the previously known protein:

Sequence name: STC2_HUMAN

Sequence documentation:

Alignment of: R20779_P2 x STC2_HUMAN ..

Alignment segment 1/1:

Quality: 1688.00

Escore: 0

Matching length: 171 Total length: 171

Matching Percent Similarity: 99.42 Matching Percent Identity: 99.42

Total Percent Similarity: 99.42 Total Percent Identity: 99.42

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNT 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNT 50
                .         .         .         .         .
     51 AEIQHCLVNAGDVGCGVFECFENNSCEIRGLHGICMTFLHNAGKFDAQGK 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 AEIQHCLVNAGDVGCGVFECFENNSCEIRGLHGICMTFLHNAGKFDAQGK 100
                .         .         .         .         .
    101 SFIKDALKCKAHALRHRFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQ 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 SFIKDALKCKAHALRHRFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQ 150
```

1178

```
                .         .         .
151 ENTRVIVEMIHFKDLLLHECY                                   171
    |||||||||||||||||||| |
151 ENTRVIVEMIHFKDLLLHEPY                                   171
```

DESCRIPTION FOR CLUSTER R38144

Cluster R38144 features 6 transcript(s) and 24 segment(s) of interest, the names for which are given in Tables 1169 and 1170, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1171.

Table 1169 - Transcripts of interest

Transcript Name	Sequence ID No.
R38144_PEA_2_T6	135
R38144_PEA_2_T10	136
R38144_PEA_2_T13	137
R38144_PEA_2_T15	138
R38144_PEA_2_T19	139
R38144_PEA_2_T27	140

Table 1170 - Segments of interest

Segment Name	Sequence ID No.
R38144_PEA_2_node_21	937
R38144_PEA_2_node_26	938
R38144_PEA_2_node_29	939
R38144_PEA_2_node_31	940
R38144_PEA_2_node_46	941
R38144_PEA_2_node_47	942
R38144_PEA_2_node_49	943
R38144_PEA_2_node_0	944

R38144_PEA_2_node_1	945
R38144_PEA_2_node_4	946
R38144_PEA_2_node_5	947
R38144_PEA_2_node_7	948
R38144_PEA_2_node_11	949
R38144_PEA_2_node_14	950
R38144_PEA_2_node_15	951
R38144_PEA_2_node_16	952
R38144_PEA_2_node_19	953
R38144_PEA_2_node_20	954
R38144_PEA_2_node_36	955
R38144_PEA_2_node_37	956
R38144_PEA_2_node_43	957
R38144_PEA_2_node_44	958
R38144_PEA_2_node_45	959
R38144_PEA_2_node_51	960

Table 1171 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
R38144_PEA_2_P6	1403	R38144_PEA_2_T6
R38144_PEA_2_P13	1404	R38144_PEA_2_T13
R38144_PEA_2_P15	1405	R38144_PEA_2_T15
R38144_PEA_2_P19	1406	R38144_PEA_2_T19
R38144_PEA_2_P24	1407	R38144_PEA_2_T27
R38144_PEA_2_P36	1408	R38144_PEA_2_T10

These sequences are variants of the known protein Putative alpha-mannosidase C20orf31 precursor (SwissProt accession identifier CT31_HUMAN; known also according to the synonyms EC 3.2.1), SEQ ID NO:1459, referred to herein as the previously known protein.

The sequence for protein Putative alpha-mannosidase C20orf31 precursor is given at the end of the application, as "Putative alpha-mannosidase C20orf31 precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1172.

Table 1172 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
456	A -> T. /FTId=VAR_012165.
511	S -> C

Protein Putative alpha-mannosidase C20orf31 precursor localization is believed to be Secreted (Potential).

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: carbohydrate metabolism; N-linked glycosylation, which are annotation(s) related to Biological Process; mannosyl-oligosaccharide 1,2-alpha-mannosidase; calcium binding; hydrolase, acting on glycosyl bonds, which are annotation(s) related to Molecular Function; and membrane, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih. gov/projects/LocusLink/>.

Cluster R38144 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 45 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 45 and Table 1173. This cluster is overexpressed (at least at a minimum level) in the

following pathological conditions: epithelial malignant tumors, lung malignant tumors, skin malignancies and gastric carcinoma.

Table 1173 - Normal tissue distribution

Name of Tissue	Number
Adrenal	40
Bladder	41
Bone	38
Brain	16
Colon	37
Epithelial	18
General	31
head and neck	50
Kidney	26
Liver	4
Lung	11
lymph nodes	47
Breast	52
Ovary	7
Pancreas	20
Prostate	0
Skin	13
Stomach	0
Uterus	0

Table 1174 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
Adrenal	9.2e-01	6.9e-01	1	0.5	7.8e-01	0.9
Bladder	7.6e-01	8.1e-01	8.1e-01	0.9	9.0e-01	0.7
Bone	6.6e-01	8.5e-01	1	0.6	1	0.6

Brain	8.0e-02	6.0e-02	4.7e-02	3.0	1.6e-02	3.0
colon	7.7e-01	7.5e-01	1	0.5	3.5e-01	0.8
epithelial	2.0e-01	4.8e-03	1.7e-01	1.4	2.7e-16	5.2
general	3.9e-01	2.2e-02	7.8e-01	0.9	2.1e-19	2.9
head and neck	3.4e-01	5.6e-01	4.6e-01	1.4	7.5e-01	0.9
kidney	8.3e-01	7.7e-01	4.4e-01	1.4	8.5e-02	1.6
liver	9.1e-01	6.0e-01	1	0.9	1.1e-01	1.8
lung	1.6e-02	1.5e-02	9.5e-02	3.8	1.6e-05	6.6
lymph nodes	7.1e-01	7.8e-01	1	0.3	1.2e-04	1.0
breast	9.1e-01	9.1e-01	1	0.5	9.7e-01	0.6
ovary	5.0e-01	2.9e-01	4.7e-01	1.7	7.0e-02	2.2
pancreas	7.2e-01	4.2e-01	8.1e-01	0.8	3.0e-02	1.8
prostate	7.9e-01	5.7e-01	3.0e-01	2.5	1.8e-04	3.0
skin	9.2e-01	8.7e-02	1	0.5	3.0e-05	4.1
stomach	3.0e-01	5.5e-02	2.5e-01	3.0	9.2e-04	6.1
uterus	2.1e-01	9.4e-02	4.4e-01	2.0	5.1e-01	1.9

As noted above, cluster R38144 features 6 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) ofprotein Putative alpha-mannosidase C20orf31 precursor. A description of each variant protein according to the present invention is now provided.

Variant protein R38144_PEA_2_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T6. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P6 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P6, comprising a first amino acid sequence being at least 90 % homologous to

MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD
ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET
NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV
NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV
GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR
FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGG
LPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESIEKISKVEC
GFAT corresponding to amino acids 1 - 412 of CT31_HUMAN, which also corresponds to amino acids 1 - 412 of R38144_PEA_2_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence LASFSHMSDQRSARPQAGQPHGVVLPGRDCEIPLPPV corresponding to amino acids 413 - 449 of R38144_PEA_2_P6, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence LASFSHMSDQRSARPQAGQPHGVVLPGRDCEIPLPPV in R38144_PEA_2_P6.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1175, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1175 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
54	A -> V	Yes
55	F -> L	Yes
73	S -> I	Yes
87	I ->	No
145	P ->	No
145	P -> A	No
164	A -> G	No
164	A ->	No
203	A -> G	No
203	A ->	No
211	D ->	No
236	G ->	No
265	V -> G	No
285	K ->	No
294	D -> N	No
305	G -> E	No
323	Q -> R	No
346	F ->	No

The glycosylation sites of variant protein R38144_PEA_2_P6, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1176 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1176 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
450	no	
289	yes	289
112	yes	112
90	yes	90

Variant protein R38144_PEA_2_P6 is encoded by the following transcript(s): R38144_PEA_2_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T6 is shown in bold; this coding portion starts at position 91 and ends at position 1437. The transcript also has the following SNPs as listed in Table 1177 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1177 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
251	C -> T	Yes
253	T -> C	Yes
308	G -> T	Yes
312	T -> C	No
350	T ->	No
523	C ->	No
523	C -> G	No
581	C ->	No
581	C -> G	No
698	C ->	No

698	C -> G	No
723	C ->	No
798	C ->	No
798	C -> G	No
849	-> C	No
849	-> G	No
884	T -> G	No
901	-> C	No
901	-> T	No
943	A ->	No
970	G -> A	No
1004	G -> A	No
1058	A -> G	No
1126	T ->	No
1218	C -> T	Yes
1392	A -> G	No
1425	T -> C	No
1481	G -> A	Yes
1560	C -> T	No
1566	C ->	No
1644	G -> A	Yes
1646	A -> T	No
1763	A ->	No
1763	A -> C	No
1781	C -> T	Yes
1799	C ->	No
1799	C -> G	No
1844	T -> G	No
1855	A -> C	Yes

Variant protein R38144_PEA_2_P13 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T13. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P13 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P13, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQ corresponding to amino acids 1 - 323 of CT31_HUMAN, which also corresponds to amino acids 1 - 323 of R38144_PEA_2_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence NLLKAQCTSTVPRGIPPS corresponding to amino acids 324 - 341 of R38144_PEA_2_P13, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence NLLKAQCTSTVPRGIPPS in R38144_PEA_2_P13.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P13 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1178, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1178 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
54	A -> V	Yes
55	F -> L	Yes
73	S -> I	Yes
87	I ->	No
145	P ->	No
145	P -> A	No
164	A -> G	No
164	A ->	No
203	A -> G	No
203	A ->	No
211	D ->	No
236	G ->	No
265	V -> G	No
285	K ->	No
294	D -> N	No
305	G -> E	No
323	Q -> R	No

328	A -> V	Yes

The glycosylation sites of variant protein R38144_PEA_2_P13, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1179 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1179 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
450	no	
289	yes	289
112	yes	112
90	yes	90

Variant protein R38144_PEA_2_P13 is encoded by the following transcript(s): R38144_PEA_2_T13, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T13 is shown in bold; this coding portion starts at position 91 and ends at position 1113. The transcript also has the following SNPs as listed in Table 1180 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1180 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
251	C -> T	Yes
253	T -> C	Yes

308	G -> T	Yes
312	T -> C	No
350	T ->	No
523	C ->	No
523	C -> G	No
581	C ->	No
581	C -> G	No
698	C ->	No
698	C -> G	No
723	C ->	No
798	C ->	No
798	C -> G	No
849	-> C	No
849	-> G	No
884	T -> G	No
901	-> C	No
901	-> T	No
943	A ->	No
970	G -> A	No
1004	G -> A	No
1058	A -> G	No
1073	C -> T	Yes
1222	A -> G	No
1255	T -> C	No
1311	G -> A	Yes
1390	C -> T	No
1396	C ->	No
1474	G -> A	Yes
1476	A -> T	No
1593	A ->	No

1593	A -> C	No
1611	C -> T	Yes
1629	C ->	No
1629	C -> G	No
1674	T -> G	No
1685	A -> C	Yes

Variant protein R38144_PEA_2_P15 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T15. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P15 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P15, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLE corresponding to amino acids 1 - 282 of CT31_HUMAN, which also corresponds to amino acids 1 - 282 of R38144_PEA_2_P15, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence PHWRH corresponding to amino acids 283 - 287 of R38144_PEA_2_P15, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P15, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence PHWRH in R38144_PEA_2_P15.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P15 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1181, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1181 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
54	A -> V	Yes
55	F -> L	Yes
73	S -> I	Yes
87	I ->	No
145	P ->	No
145	P -> A	No
164	A -> G	No
164	A ->	No
203	A -> G	No
203	A ->	No
211	D ->	No

236	G ->	No
265	V -> G	No

The glycosylation sites of variant protein R38144_PEA_2_P15, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1182 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1182 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
450	no	
289	no	
112	yes	112
90	yes	90

Variant protein R38144_PEA_2_P15 is encoded by the following transcript(s): R38144_PEA_2_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T15 is shown in bold; this coding portion starts at position 91 and ends at position 951. The transcript also has the following SNPs as listed in Table 1183 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P15 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1183- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
251	C -> T	Yes

253	T -> C	Yes
308	G -> T	Yes
312	T -> C	No
350	T ->	No
523	C ->	No
523	C -> G	No
581	C ->	No
581	C -> G	No
698	C ->	No
698	C -> G	No
723	C ->	No
798	C ->	No
798	C -> G	No
849	-> C	No
849	-> G	No
884	T -> G	No
901	-> C	No
901	-> T	No
1001	T ->	No
1093	C -> T	Yes
1242	A -> G	No
1275	T -> C	No
1331	G -> A	Yes
1410	C -> T	No
1416	C ->	No
1494	G -> A	Yes
1496	A -> T	No
1613	A ->	No
1613	A -> C	No
1631	C -> T	Yes

1649	C ->	No
1649	C -> G	No
1694	T -> G	No
1705	A -> C	Yes

Variant protein R38144_PEA_2_P19 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T19. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P19 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P19, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTV NLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLV GNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTR FDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGG LPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESIEKISKVEC GFAT corresponding to amino acids 1 - 412 of CT31_HUMAN, which also corresponds to amino acids 1 - 412 of R38144_PEA_2_P19, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence KRSRSVAQAGVQWCDHDSPQP corresponding to amino acids 413 - 433 of R38144_PEA_2_P19, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P19, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence KRSRSVAQAGVQWCDHDSPQP in R38144_PEA_2_P19.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P19 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1184, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1184- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
54	A -> V	Yes
55	F -> L	Yes
73	S -> I	Yes
87	I ->	No
145	P ->	No
145	P -> A	No
164	A -> G	No
164	A ->	No
203	A -> G	No

203	A ->	No
211	D ->	No
236	G ->	No
265	V -> G	No
285	K ->	No
294	D -> N	No
305	G -> E	No
323	Q -> R	No
346	F ->	No

The glycosylation sites of variant protein R38144_PEA_2_P19, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1185(given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1185- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
450	no	
289	yes	289
112	yes	112
90	yes	90

Variant protein R38144_PEA_2_P19 is encoded by the following transcript(s): R38144_PEA_2_T19, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T19 is shown in bold; this coding portion starts at position 91 and ends at position 1389. The transcript also has the following SNPs as listed in Table 1186 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein R38144_PEA_2_P19 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1186- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
251	C -> T	Yes
253	T -> C	Yes
308	G -> T	Yes
312	T -> C	No
350	T ->	No
523	C ->	No
523	C -> G	No
581	C ->	No
581	C -> G	No
698	C ->	No
698	C -> G	No
723	C ->	No
798	C ->	No
798	C -> G	No
849	-> C	No
849	-> G	No
884	T -> G	No
901	-> C	No
901	-> T	No
943	A ->	No
970	G -> A	No
1004	G -> A	No
1058	A -> G	No

1126	T ->	No
1218	C -> T	Yes
1446	C ->	Yes

Variant protein R38144_PEA_2_P24 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T27. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P24 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P24, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFD ELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFET NIR corresponding to amino acids 1 - 121 of CT31_HUMAN, which also corresponds to amino acids 1 - 121 of R38144_PEA_2_P24, and a second amino acid sequence being at least 90 % homologous to EYNKAIRNYTRFDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLN YYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDA VESIEKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNNGSTFDA VITPYGECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVEDLMREFYSLKRSRSKFQ KNTVSSGPWEPPARPGTLFSPENHDQARERKPAKQKVPLLSCPSQPFTSKLALLGQVFL DSS corresponding to amino acids 282 - 578 of CT31_HUMAN, which also corresponds to amino acids 122 - 418 of R38144_PEA_2_P24, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated chimeric polypeptide encoding for an edge portion of R38144_PEA_2_P24, comprising a polypeptide having a length "n", wherein n is at least about 10 amino acids in length, optionally at least about 20 amino acids in length, preferably at least about 30 amino

acids in length, more preferably at least about 40 amino acids in length and most preferably at least about 50 amino acids in length, wherein at least two amino acids comprise RE, having a structure as follows: a sequence starting from any of amino acid numbers 121-x to 121; and ending at any of amino acid numbers 122+ ((n-2) - x), in which x varies from 0 to n-2.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P24 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1187, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P24 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1187- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
54	A -> V	Yes
55	F -> L	Yes
73	S -> I	Yes
87	I ->	No
125	K ->	No
134	D -> N	No
145	G -> E	No
163	Q -> R	No
186	F ->	No

266	E -> G	No
277	L -> P	No
296	A -> T	Yes
322	P -> L	No
324	A ->	No
350	R -> Q	Yes
351	S -> C	No
390	K ->	No
390	K -> Q	No
396	L -> F	Yes
402	P ->	No
402	P -> A	No
417	S -> A	No

The glycosylation sites of variant protein R38144_PEA_2_P24, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1188 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1188- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
450	yes	290
289	yes	129
112	yes	112
90	yes	90

Variant protein R38144_PEA_2_P24 is encoded by the following transcript(s): R38144_PEA_2_T27, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T27 is shown in bold; this coding portion starts at

position 91 and ends at position 1344. The transcript also has the following SNPs as listed in Table 1189 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P24 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1189 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
251	C -> T	Yes
253	T -> C	Yes
308	G -> T	Yes
312	T -> C	No
350	T ->	No
463	A ->	No
490	G -> A	No
524	G -> A	No
578	A -> G	No
646	T ->	No
738	C -> T	Yes
887	A -> G	No
920	T -> C	No
976	G -> A	Yes
1055	C -> T	No
1061	C ->	No
1139	G -> A	Yes
1141	A -> T	No
1258	A ->	No
1258	A -> C	No

1276	C -> T	Yes
1294	C ->	No
1294	C -> G	No
1339	T -> G	No
1350	A -> C	Yes

Variant protein R38144_PEA_2_P36 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R38144_PEA_2_T10. An alignment is given to the known protein (Putative alpha-mannosidase C20orf31 precursor; SEQ ID NO:1459) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R38144_PEA_2_P36 and AAH16184 (SEQ ID NO: 1460):

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR corresponding to amino acids 1 - 36 of AAH16184, which also corresponds to amino acids 1 - 36 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence FWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 37 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence FWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

Comparison report between R38144_PEA_2_P36 and AAQ88943 (SEQ ID NO:1461):

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHY corresponding to amino acids 1 - 35 of AAQ88943, which also corresponds to amino acids 1 - 35 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence RFWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 36 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence RFWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

Comparison report between R38144_PEA_2_P36 and CT31_HUMAN:

1.An isolated chimeric polypeptide encoding for R38144_PEA_2_P36, comprising a first amino acid sequence being at least 90 % homologous to MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR corresponding to amino acids 1 - 36 of CT31_HUMAN, which also corresponds to amino acids 1 - 36 of R38144_PEA_2_P36, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence FWGMSQNSKEWLKCSRTAWTLILM corresponding to amino acids 37 - 60 of R38144_PEA_2_P36, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R38144_PEA_2_P36, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence FWGMSQNSKEWLKCSRTAWTLILM in R38144_PEA_2_P36.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized

programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R38144_PEA_2_P36 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1190, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P36 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1190 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
10	G ->	No
37	F ->	No

The glycosylation sites of variant protein R38144_PEA_2_P36, as compared to the known protein Putative alpha-mannosidase C20orf31 precursor, are described in Table 1191(given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1191 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
450	no
289	no
112	no
90	no

Variant protein R38144_PEA_2_P36 is encoded by the following transcript(s): R38144_PEA_2_T10, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R38144_PEA_2_T10 is shown in bold; this coding portion starts at position 91 and ends at position 270. The transcript also has the following SNPs as listed in Table 1192 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R38144_PEA_2_P36 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1192- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
120	C ->	No
199	T ->	No
372	C ->	No
372	C -> G	No
430	C ->	No
430	C -> G	No
547	C ->	No
547	C -> G	No
572	C ->	No
647	C ->	No
647	C -> G	No
698	-> C	No
698	-> G	No
733	T -> G	No
750	-> C	No
750	-> T	No
792	A ->	No
819	G -> A	No
853	G -> A	No

907	A -> G	No
975	T ->	No
1067	C -> T	Yes
1216	A -> G	No
1249	T -> C	No
1305	G -> A	Yes
1384	C -> T	No
1390	C ->	No
1468	G -> A	Yes
1470	A -> T	No
1587	A ->	No
1587	A -> C	No
1605	C -> T	Yes
1623	C ->	No
1623	C -> G	No
1668	T -> G	No
1679	A -> C	Yes

As noted above, cluster R38144 features 24 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R38144_PEA_2_node_21 according to the present invention is supported by 108 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1193 below describes the starting and ending position of this segment on each transcript.

Table 1193- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	626	792
R38144_PEA_2_T10	475	641
R38144_PEA_2_T13	626	792
R38144_PEA_2_T15	626	792
R38144_PEA_2_T19	626	792

Segment cluster R38144_PEA_2_node_26 according to the present invention is supported by 98 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1194 below describes the starting and ending position of this segment on each transcript.

Table 1194- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	793	934
R38144_PEA_2_T10	642	783
R38144_PEA_2_T13	793	934
R38144_PEA_2_T15	793	934
R38144_PEA_2_T19	793	934

Segment cluster R38144_PEA_2_node_29 according to the present invention is supported by 98 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1195 below describes the starting and ending position of this segment on each transcript.

Table 1195- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	935	1059
R38144_PEA_2_T10	784	908
R38144_PEA_2_T13	935	1059
R38144_PEA_2_T19	935	1059
R38144_PEA_2_T27	455	579

Segment cluster R38144_PEA_2_node_31 according to the present invention is supported by 95 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1196 below describes the starting and ending position of this segment on each transcript.

Table 1196 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1060	1204
R38144_PEA_2_T10	909	1053
R38144_PEA_2_T15	935	1079
R38144_PEA_2_T19	1060	1204
R38144_PEA_2_T27	580	724

Segment cluster R38144_PEA_2_node_46 according to the present invention is supported by 147 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T27. Table 1197 below describes the starting and ending position of this segment on each transcript.

Table 1197- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1373	1544
R38144_PEA_2_T10	1197	1368
R38144_PEA_2_T13	1203	1374
R38144_PEA_2_T15	1223	1394
R38144_PEA_2_T27	868	1039

Segment cluster R38144_PEA_2_node_47 according to the present invention is supported by 147 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T27. Table 1198 below describes the starting and ending position of this segment on each transcript.

Table 1198- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1545	1919
R38144_PEA_2_T10	1369	1743
R38144_PEA_2_T13	1375	1749
R38144_PEA_2_T15	1395	1769
R38144_PEA_2_T27	1040	1414

Segment cluster R38144_PEA_2_node_49 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T19. Table 1199 below describes the starting and ending position of this segment on each transcript.

Table 1199- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T19	1327	1448

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R38144_PEA_2_node_0 according to the present invention is supported by 101 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1200 below describes the starting and ending position of this segment on each transcript.

Table 1201- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1	105
R38144_PEA_2_T10	1	105
R38144_PEA_2_T13	1	105
R38144_PEA_2_T15	1	105
R38144_PEA_2_T19	1	105
R38144_PEA_2_T27	1	105

Segment cluster R38144_PEA_2_node_1 according to the present invention is supported by 105 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1202 below describes the starting and ending position of this segment on each transcript.

Table 1202- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	106	197
R38144_PEA_2_T10	106	197
R38144_PEA_2_T13	106	197
R38144_PEA_2_T15	106	197
R38144_PEA_2_T19	106	197
R38144_PEA_2_T27	106	197

Segment cluster R38144_PEA_2_node_4 according to the present invention is supported by 107 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1203 below describes the starting and ending position of this segment on each transcript.

Table 1203- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	198	299
R38144_PEA_2_T13	198	299
R38144_PEA_2_T15	198	299
R38144_PEA_2_T19	198	299
R38144_PEA_2_T27	198	299

Segment cluster R38144_PEA_2_node_5 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1204 below describes the starting and ending position of this segment on each transcript.

Table 1204- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	300	308
R38144_PEA_2_T13	300	308
R38144_PEA_2_T15	300	308
R38144_PEA_2_T19	300	308
R38144_PEA_2_T27	300	308

Segment cluster R38144_PEA_2_node_7 according to the present invention is supported by 92 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1205 below describes the starting and ending position of this segment on each transcript.

Table 1205- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	309	348
R38144_PEA_2_T13	309	348
R38144_PEA_2_T15	309	348
R38144_PEA_2_T19	309	348
R38144_PEA_2_T27	309	348

Segment cluster R38144_PEA_2_node_11 according to the present invention is supported by 106 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1206 below describes the starting and ending position of this segment on each transcript.

Table 1206- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	349	454
R38144_PEA_2_T10	198	303
R38144_PEA_2_T13	349	454
R38144_PEA_2_T15	349	454
R38144_PEA_2_T19	349	454
R38144_PEA_2_T27	349	454

Segment cluster R38144_PEA_2_node_14 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1207 below describes the starting and ending position of this segment on each transcript.

Table 1207- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	455	460
R38144_PEA_2_T10	304	309
R38144_PEA_2_T13	455	460
R38144_PEA_2_T15	455	460
R38144_PEA_2_T19	455	460

Segment cluster R38144_PEA_2_node_15 according to the present invention is supported by 105 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1208 below describes the starting and ending position of this segment on each transcript.

Table 1208- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	461	487
R38144_PEA_2_T10	310	336
R38144_PEA_2_T13	461	487
R38144_PEA_2_T15	461	487
R38144_PEA_2_T19	461	487

Segment cluster R38144_PEA_2_node_16 according to the present invention is supported by 106 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1209 below describes the starting and ending position of this segment on each transcript.

Table 1209- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	488	580
R38144_PEA_2_T10	337	429
R38144_PEA_2_T13	488	580
R38144_PEA_2_T15	488	580
R38144_PEA_2_T19	488	580

Segment cluster R38144_PEA_2_node_19 according to the present invention is supported by 93 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1210 below describes the starting and ending position of this segment on each transcript.

Table 1210- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	581	615
R38144_PEA_2_T10	430	464
R38144_PEA_2_T13	581	615
R38144_PEA_2_T15	581	615
R38144_PEA_2_T19	581	615

Segment cluster R38144_PEA_2_node_20 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T19. Table 1211 below describes the starting and ending position of this segment on each transcript.

Table 1211- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	616	625
R38144_PEA_2_T10	465	474
R38144_PEA_2_T13	616	625
R38144_PEA_2_T15	616	625
R38144_PEA_2_T19	616	625

Segment cluster R38144_PEA_2_node_36 according to the present invention is supported by 95 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1212 below describes the starting and ending position of this segment on each transcript.

Table 1212- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1205	1293
R38144_PEA_2_T10	1054	1142
R38144_PEA_2_T13	1060	1148
R38144_PEA_2_T15	1080	1168
R38144_PEA_2_T19	1205	1293
R38144_PEA_2_T27	725	813

Segment cluster R38144_PEA_2_node_37 according to the present invention is supported by 97 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15, R38144_PEA_2_T19 and R38144_PEA_2_T27. Table 1213 below describes the starting and ending position of this segment on each transcript.

Table 1213- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1294	1326
R38144_PEA_2_T10	1143	1175
R38144_PEA_2_T13	1149	1181
R38144_PEA_2_T15	1169	1201
R38144_PEA_2_T19	1294	1326
R38144_PEA_2_T27	814	846

Segment cluster R38144_PEA_2_node_43 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6. Table 1214 below describes the starting and ending position of this segment on each transcript.

Table 1214- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1327	1346

Segment cluster R38144_PEA_2_node_44 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6. Table 1215 below describes the starting and ending position of this segment on each transcript.

Table 1215- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1347	1351

Segment cluster R38144_PEA_2_node_45 according to the present invention can be found in the following transcript(s): R38144_PEA_2_T6, R38144_PEA_2_T10, R38144_PEA_2_T13, R38144_PEA_2_T15 and R38144_PEA_2_T27. Table 1216 below describes the starting and ending position of this segment on each transcript.

Table 1216- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T6	1352	1372
R38144_PEA_2_T10	1176	1196
R38144_PEA_2_T13	1182	1202
R38144_PEA_2_T15	1202	1222
R38144_PEA_2_T27	847	867

Segment cluster R38144_PEA_2_node_51 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R38144_PEA_2_T19. Table 1217 below describes the starting and ending position of this segment on each transcript.

Table 1217 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R38144_PEA_2_T19	1449	1522

Variant protein alignment to the previously known protein:

```
Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P6 x CT31_HUMAN    ..

Alignment segment 1/1:

                        Quality: 4031.00
Escore:        0
                Matching length:      413                  Total
length:      413
 Matching Percent Similarity:  100.00   Matching Percent
Identity:    99.76
```

1220

Total Percent Similarity: 100.00 Total Percent Identity: 99.76

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
                .         .         .         .         .
 51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
                .         .         .         .         .
101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
                .         .         .         .         .
151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
                .         .         .         .         .
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
                .         .         .         .         .
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
                .         .         .         .         .
301 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 350
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        301 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 350
                    .         .         .         .         .
        351 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        351 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 400
                    .
        401 EKISKVECGFATL                                      413
            ||||||||||||:
        401 EKISKVECGFATI                                      413
```

Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P13 x CT31_HUMAN ..

Alignment segment 1/1:

Quality: 3167.00

Escore: 0

Matching length: 326 Total length: 326

Matching Percent Similarity: 100.00 Matching Percent Identity: 99.39

1222

Total Percent Similarity: 100.00 Total Percent Identity: 99.39

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
                .         .         .         .         .
 51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
                .         .         .         .         .
101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
                .         .         .         .         .
151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
                .         .         .         .         .
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
                .         .         .         .         .
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
                .         .
301 QMYKGTVSMPVFQSLEAYWPGLQNLL                          326
```

1223

||||||||||||||||||||||:|:

301 QMYKGTVSMPVFQSLEAYWPGLQSLI 326

Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P15 x CT31_HUMAN ..

Alignment segment 1/1:

Quality: 2725.00

Escore: 0

Matching length: 282 Total length: 282

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

. . . . .

1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50

||||||||||||||||||||||||||||||||||||||||||||||||||

1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50

```
                .         .         .         .         .
        51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||
        51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
                .         .         .         .         .
       101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
           |||||||||||||||||||||||||||||||||||||||||||||||||||
       101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
                .         .         .         .         .
       151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
           |||||||||||||||||||||||||||||||||||||||||||||||||||
       151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
                .         .         .         .         .
       201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
           |||||||||||||||||||||||||||||||||||||||||||||||||||
       201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
                .         .         .
       251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLE                   282
           ||||||||||||||||||||||||||||||
       251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLE                   282
```

Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P19 x CT31_HUMAN ..

Alignment segment 1/1:

Quality: 4029.00
Escore: 0
Matching length: 412 Total length: 412
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
                .         .         .         .         .
     51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
                .         .         .         .         .
    101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
                .         .         .         .         .
    151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
                .         .         .         .         .
```

```
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
          .         .         .         .         .
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
          .         .         .         .         .
301 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 350
          .         .         .         .         .
351 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 400
          .
401 EKISKVECGFAT                                       412
    ||||||||||||
401 EKISKVECGFAT                                       412
```

Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P24 x CT31_HUMAN ..

Alignment segment 1/1:

Quality: 4063.00

Escore: 0

Matching length: 418 Total length: 578

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 72.32 Total Percent Identity: 72.32

Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSY 50
                .         .         .         .         .
     51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 LENAFPFDELRPLTCDGHDTWGSFSLTLIDALDTLLILGNVSEFQRVVEV 100
                .         .         .         .         .
    101 LQDSVDFDIDVNASVFETNIR..........................  121
        |||||||||||||||||||||
    101 LQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPL 150
                .         .         .         .         .
    121 ..................................................  121

    151 LRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIV 200
                .         .         .         .         .
    121 ..................................................  121
```

```
201 EFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQD 250
                .         .         .         .         .
122 ..............................EYNKAIRNYTRFDDWYLWV 140
                                    ||||||||||||||||||
251 AGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWV 300
                .         .         .         .         .
141 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 190
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 QMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLP 350
                .         .         .         .         .
191 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 240
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
351 EFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESI 400
                .         .         .         .         .
241 EKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNN 290
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
401 EKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNN 450
                .         .         .         .         .
291 GSTFDAVITPYGECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVED 340
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
451 GSTFDAVITPYGECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVED 500
                .         .         .         .         .
341 LMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPENHDQARERKPAK 390
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
501 LMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPENHDQARERKPAK 550
                .         .
391 QKVPLLSCPSQPFTSKLALLGQVFLDSS                         418
    ||||||||||||||||||||||||||||
551 QKVPLLSCPSQPFTSKLALLGQVFLDSS                         578
```

Sequence name: AAH16184

Sequence documentation:

Alignment of: R38144_PEA_2_P36 x AAH16184 ..

Alignment segment 1/1:

Quality: 364.00
Escore: 0
Matching length: 36 Total length: 36
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR              36
          ||||||||||||||||||||||||||||||||||||
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR              36
```

Sequence name: AAQ88943

Sequence documentation:

Alignment of: R38144_PEA_2_P36 x AAQ88943 ..

Alignment segment 1/1:

Quality: 362.00
Escore: 0
Matching length: 37 Total length: 37
Matching Percent Similarity: 97.30 Matching Percent Identity: 97.30
Total Percent Similarity: 97.30 Total Percent Identity: 97.30
Gaps: 0

Alignment:

```
                .         .         .
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRF  37
          |||||||||||||||||||||||||||||||||||| |
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYSF  37
```

```
Sequence name: CT31_HUMAN

Sequence documentation:

Alignment of: R38144_PEA_2_P36 x CT31_HUMAN    ..

Alignment segment 1/1:

                      Quality:  364.00
Escore:        0
              Matching length:       36                 Total
length:       36
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                         Gaps:        0

Alignment:
                    .         .         .
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR                36
          ||||||||||||||||||||||||||||||||||||
        1 MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYR                36
```

DESCRIPTION FOR CLUSTER HUMOSTRO

Cluster HUMOSTRO features 3 transcript(s) and 30 segment(s) of interest, the names for which are given in Tables 1218 and 1219, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1220.

Table 1218 - Transcripts of interest

Transcript Name	Sequence ID No.
HUMOSTRO_PEA_1_PEA_1_T14	141
HUMOSTRO_PEA_1_PEA_1_T16	142
HUMOSTRO_PEA_1_PEA_1_T30	143

Table 1219- Segments of interest

Segment Name	Sequence ID No.
HUMOSTRO_PEA_1_PEA_1_node_0	961
HUMOSTRO_PEA_1_PEA_1_node_10	962
HUMOSTRO_PEA_1_PEA_1_node_16	963
HUMOSTRO_PEA_1_PEA_1_node_23	964
HUMOSTRO_PEA_1_PEA_1_node_31	965
HUMOSTRO_PEA_1_PEA_1_node_43	966
HUMOSTRO_PEA_1_PEA_1_node_3	967
HUMOSTRO_PEA_1_PEA_1_node_5	968
HUMOSTRO_PEA_1_PEA_1_node_7	969
HUMOSTRO_PEA_1_PEA_1_node_8	970
HUMOSTRO_PEA_1_PEA_1_node_15	971
HUMOSTRO_PEA_1_PEA_1_node_17	972
HUMOSTRO_PEA_1_PEA_1_node_20	973
HUMOSTRO_PEA_1_PEA_1_node_21	974
HUMOSTRO_PEA_1_PEA_1_node_22	975
HUMOSTRO_PEA_1_PEA_1_node_24	976
HUMOSTRO_PEA_1_PEA_1_node_26	977
HUMOSTRO_PEA_1_PEA_1_node_27	978
HUMOSTRO_PEA_1_PEA_1_node_28	979
HUMOSTRO_PEA_1_PEA_1_node_29	980
HUMOSTRO_PEA_1_PEA_1_node_30	981
HUMOSTRO_PEA_1_PEA_1_node_32	982

HUMOSTRO_PEA_1_PEA_1_node_34	983
HUMOSTRO_PEA_1_PEA_1_node_36	984
HUMOSTRO_PEA_1_PEA_1_node_37	985
HUMOSTRO_PEA_1_PEA_1_node_38	986
HUMOSTRO_PEA_1_PEA_1_node_39	987
HUMOSTRO_PEA_1_PEA_1_node_40	988
HUMOSTRO_PEA_1_PEA_1_node_41	989
HUMOSTRO_PEA_1_PEA_1_node_42	990

Table 1220- Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
HUMOSTRO_PEA_1_PEA_1_P21	1627	HUMOSTRO_PEA_1_PEA_1_T14
HUMOSTRO_PEA_1_PEA_1_P25	1628	HUMOSTRO_PEA_1_PEA_1_T16
HUMOSTRO_PEA_1_PEA_1_P30	1629	HUMOSTRO_PEA_1_PEA_1_T30

These sequences are variants of the known protein Osteopontin precursor (SwissProt accession identifier OSTP_HUMAN; known also according to the synonyms Bone sialoprotein 1; Urinary stone protein; Secreted phosphoprotein 1; SPP-1; Nephropontin; Uropontin), SEQ ID NO:1462, referred to herein as the previously known protein.

Protein Osteopontin precursor is known or believed to have the following function(s): Binds tightly to hydroxyapatite. Appears to form an integral part of the mineralized matrix. Probably important to cell-matrix interaction. Acts as a cytokine involved in enhancing production of interferon-gamma and interleukin-12 and reducing production of interleukin-10 and is essential in the pathway that leads to type I immunity (By similarity). The sequence for protein Osteopontin precursor is given at the end of the application, as "Osteopontin precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1221.

Table 1221- Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
301	R -> H (in dbSNP:4660). /FTId=VAR_014717.
188	D -> H
237	T -> A
275 – 278	SHEF -> GNSL

Protein Osteopontin precursor localization is believed to be Secreted.

The previously known protein also has the following indication(s) and/or potential therapeutic use(s): Regeneration, bone. It has been investigated for clinical/therapeutic use in humans, for example as a target for an antibody or small molecule, and/or as a direct therapeutic; available information related to these investigations is as follows. Potential pharmaceutically related or therapeutically related activity or activities of the previously known protein are as follows: Bone formation stimulant. A therapeutic role for a protein represented by the cluster has been predicted. The cluster was assigned this field because there was information in the drug database or the public databases (e.g., described herein above) that this protein, or part thereof, is used or can be used for a potential therapeutic indication: Musculoskeletal.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: ossification; anti-apoptosis; inflammatory response; cell-matrix adhesion; cell-cell signaling, which are annotation(s) related to Biological Process; defense/immunity protein; cytokine; integrin ligand; protein binding; growth factor; apoptosis inhibitor, which are annotation(s) related to Molecular Function; and extracellular matrix, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster HUMOSTRO can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of

the table and the numbers on the y-axis of figure 46 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 46 and Table 1222. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues, lung malignant tumors, breast malignant tumors, ovarian carcinoma and skin malignancies.

Table 1222- Normal tissue distribution

Name of Tissue	Number
Adrenal	4
Bladder	0
Bone	897
Brain	506
Colon	69
Epithelial	548
General	484
head and neck	50
Kidney	5618
Liver	4
Lung	10
lymph nodes	75
Breast	8
bone marrow	62
Muscle	37
Ovary	40
Pancreas	845
Prostate	48

Skin	13
Stomach	73
Thyroid	0
Uterus	168

Table 1223- P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
Adrenal	1.5e-01	2.1e-01	2.0e-02	4.6	4.4e-02	3.6
Bladder	1.2e-01	9.2e-02	5.7e-02	4.1	2.1e-02	4.3
Bone	4.9e-01	7.4e-01	4.1e-06	0.6	5.4e-01	0.4
Brain	6.6e-01	7.0e-01	3.2e-01	0.6	1	0.4
Colon	2.7e-01	4.0e-01	3.1e-01	1.5	5.2e-01	1.1
Epithelial	2.0e-07	1.6e-03	9.8e-01	0.7	1	0.5
General	1.2e-06	1.2e-02	7.9e-01	0.8	1	0.6
head and neck	3.4e-01	5.0e-01	1	0.7	1	0.7
Kidney	6.8e-01	7.4e-01	1	0.2	1	0.1
Liver	3.3e-01	2.5e-01	1	1.8	2.3e-01	2.6
Lung	4.3e-04	4.6e-03	2.1e-30	15.0	2.8e-27	23.5
lymph nodes	6.7e-01	8.7e-01	8.1e-01	0.7	9.9e-01	0.3
Breast	2.3e-01	3.0e-01	1.9e-04	6.2	4.1e-03	4.3
bone marrow	7.5e-01	7.8e-01	1	0.3	2.0e-02	1.2
Muscle	4.0e-02	7.5e-02	1.1e-01	4.6	5.1e-01	1.5
Ovary	4.7e-02	8.4e-02	1.9e-05	5.4	8.3e-04	3.7
Pancreas	5.0e-02	3.3e-01	1	0.3	1	0.2
Prostate	8.5e-01	9.0e-01	8.9e-01	0.7	9.5e-01	0.6
Skin	1.6e-01	1.6^{e}-01	1.2e-10	12.6	5.2e-04	4.1
Stomach	1.5e-01	6.3^{e}-01	5.0e-01	1.2	9.4e-01	0.6
Thyroid	2.9e-01	2.9e-01	5.9e-02	2.0	5.9e-02	2.0
Uterus	6.1e-02	5.7^{e}-01	1.1e-01	1.3	7.0e-01	0.7

As noted above, cluster HUMOSTRO features 3 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein Osteopontin precursor. A description of each variant protein according to the present invention is now provided.

Variant protein HUMOSTRO_PEA_1_PEA_1_P21 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMOSTRO_PEA_1_PEA_1_T14. An alignment is given to the known protein (Osteopontin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMOSTRO_PEA_1_PEA_1_P21 and OSTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P21, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQKQNLLAPQ corresponding to amino acids 1 - 58 of OSTP_HUMAN, which also corresponds to amino acids 1 - 58 of HUMOSTRO_PEA_1_PEA_1_P21, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence VFLNFS corresponding to amino acids 59 - 64 of HUMOSTRO_PEA_1_PEA_1_P21, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMOSTRO_PEA_1_PEA_1_P21, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VFLNFS in HUMOSTRO_PEA_1_PEA_1_P21.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because of manual inspection of known protein localization and/or gene structure.

Variant protein HUMOSTRO_PEA_1_PEA_1_P21 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1224, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P21 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1224- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
7	C -> W	No
31	Q -> R	No
47	D -> V	Yes
49	S -> P	No

The glycosylation sites of variant protein HUMOSTRO_PEA_1_PEA_1_P21, as compared to the known protein Osteopontin precursor, are described in Table 1225 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1225- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
79	no
106	no

Variant protein HUMOSTRO_PEA_1_PEA_1_P21 is encoded by the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMOSTRO_PEA_1_PEA_1_T14 is

shown in bold; this coding portion starts at position 199 and ends at position 390. The transcript also has the following SNPs as listed in Table 1226 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P21 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1226- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	A -> G	Yes
154	T ->	No
159	G -> T	Yes
219	C -> G	No
274	-> G	No
290	A -> G	No
338	A -> T	Yes
343	T -> C	No
413	G -> C	Yes
707	C -> T	Yes
708	C -> A	Yes
715	A -> G	Yes
730	A -> C	No
730	A -> G	No
746	T -> C	Yes
767	C -> T	No
779	G -> A	Yes
866	-> G	No
869	T ->	No
889	-> A	No
891	A -> C	No

891	A -> G	No
905	T -> C	No
910	-> G	No
910	-> T	No
997	A -> G	No
1026	G -> C	No
1042	-> G	No
1042	-> T	No
1071	A ->	No
1071	A -> C	No
1098	A ->	No
1105	C -> T	No
1124	-> G	No
1135	G -> A	Yes
1136	T ->	No
1136	T -> G	No
1173	A -> C	No
1173	A -> G	No
1179	A -> G	No
1214	C -> T	Yes
1246	T ->	No
1246	T -> A	No
1359	A ->	No
1359	A -> G	No
1362	T ->	No
1365	C -> T	Yes
1366	G -> A	Yes
1408	A -> C	No
1418	A -> C	No
1433	A -> C	No

1456	A -> C	No
1524	T -> A	No
1524	T -> C	No
1547	A -> G	Yes
1553	T ->	No
1574	-> G	No
1654	A -> C	Yes
1691	A -> G	No
1703	A -> C	Yes
1755	A -> C	No
1764	T ->	No

Variant protein HUMOSTRO_PEA_1_PEA_1_P25 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMOSTRO_PEA_1_PEA_1_T16. An alignment is given to the known protein (Osteopontin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMOSTRO_PEA_1_PEA_1_P25 and OSTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P25, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQ corresponding to amino acids 1 - 31 of OSTP_HUMAN, which also corresponds to amino acids 1 - 31 of HUMOSTRO_PEA_1_PEA_1_P25, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence H corresponding to amino acids 32 - 32 of HUMOSTRO_PEA_1_PEA_1_P25, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMOSTRO_PEA_1_PEA_1_P25 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1227, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P25 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1227- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
7	C -> W	No
31	Q -> R	No

The glycosylation sites of variant protein HUMOSTRO_PEA_1_PEA_1_P25, as compared to the known protein Osteopontin precursor, are described in Table 1228 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1228- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
79	no
106	no

Variant protein HUMOSTRO_PEA_1_PEA_1_P25 is encoded by the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMOSTRO_PEA_1_PEA_1_T16 is shown in bold; this coding portion starts at position 199 and ends at position 294. The transcript also has the following SNPs as listed in Table 1229 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P25 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1229- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	A -> G	Yes
154	T ->	No
159	G -> T	Yes
219	C -> G	No
274	-> G	No
290	A -> G	No
419	C -> T	Yes
454	G -> C	Yes
527	A -> T	Yes
532	T -> C	No
630	C -> T	Yes
631	C -> A	Yes
638	A -> G	Yes
653	A -> C	No
653	A -> G	No
669	T -> C	Yes
690	C -> T	No
702	G -> A	Yes

789	-> G	No
792	T ->	No
812	-> A	No
814	A -> C	No
814	A -> G	No
828	T -> C	No
833	-> G	No
833	-> T	No
920	A -> G	No
949	G -> C	No
965	-> G	No
965	-> T	No
994	A ->	No
994	A -> C	No
1021	A ->	No
1028	C -> T	No
1047	-> G	No
1058	G -> A	Yes
1059	T ->	No
1059	T -> G	No
1096	A -> C	No
1096	A -> G	No
1102	A -> G	No
1137	C -> T	Yes
1169	T ->	No
1169	T -> A	No
1282	A ->	No
1282	A -> G	No
1285	T ->	No
1288	C -> T	Yes

1289	G -> A	Yes
1331	A -> C	No
1341	A -> C	No
1356	A -> C	No
1379	A -> C	No
1447	T -> A	No
1447	T -> C	No
1470	A -> G	Yes
1476	T ->	No
1497	-> G	No
1577	A -> C	Yes
1614	A -> G	No
1626	A -> C	Yes
1678	A -> C	No
1687	T ->	No

Variant protein HUMOSTRO_PEA_1_PEA_1_P30 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) HUMOSTRO_PEA_1_PEA_1_T30. An alignment is given to the known protein (Osteopontin precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between HUMOSTRO_PEA_1_PEA_1_P30 and OSTP_HUMAN:

1.An isolated chimeric polypeptide encoding for HUMOSTRO_PEA_1_PEA_1_P30, comprising a first amino acid sequence being at least 90 % homologous to MRIAVICFCLLGITCAIPVKQADSGSSEEKQ corresponding to amino acids 1 - 31 of OSTP_HUMAN, which also corresponds to amino acids 1 - 31 of HUMOSTRO_PEA_1_PEA_1_P30, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most

preferably at least 95% homologous to a polypeptide having the sequence VSIFYVFI corresponding to amino acids 32 - 39 of HUMOSTRO_PEA_1_PEA_1_P30, wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of HUMOSTRO_PEA_1_PEA_1_P30, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence VSIFYVFI in HUMOSTRO_PEA_1_PEA_1_P30.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein HUMOSTRO_PEA_1_PEA_1_P30 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1230, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P30 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1230- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
7	C -> W	No
31	Q -> R	No

The glycosylation sites of variant protein HUMOSTRO_PEA_1_PEA_1_P30, as compared to the known protein Osteopontin precursor, are described in Table 1231 (given according to their position(s) on the amino acid sequence in the first column; the second column

indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1231- Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?
79	no
106	no

Variant protein HUMOSTRO_PEA_1_PEA_1_P30 is encoded by the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T30, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript HUMOSTRO_PEA_1_PEA_1_T30 is shown in bold; this coding portion starts at position 199 and ends at position 315. The transcript also has the following SNPs as listed in Table 1232 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein HUMOSTRO_PEA_1_PEA_1_P30 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1232- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
136	A -> G	Yes
154	T ->	No
159	G -> T	Yes
219	C -> G	No
274	-> G	No
290	A -> G	No

As noted above, cluster HUMOSTRO features 30 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are

of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_0 according to the present invention is supported by 333 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14, HUMOSTRO_PEA_1_PEA_1_T16 and HUMOSTRO_PEA_1_PEA_1_T30. Table 1233below describes the starting and ending position of this segment on each transcript.

Table 1234- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1	184
HUMOSTRO_PEA_1_PEA_1_T16	1	184
HUMOSTRO_PEA_1_PEA_1_T30	1	184

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_10 according to the present invention is supported by 4 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T16. Table 1235 below describes the starting and ending position of this segment on each transcript.

Table 1235- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T16	292	480

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_16 according to the present invention is supported by 6 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s):

HUMOSTRO_PEA_1_PEA_1_T14. Table 1236 below describes the starting and ending position of this segment on each transcript.

Table 1236- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	373	638

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_23 according to the present invention is supported by 334 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1237 below describes the starting and ending position of this segment on each transcript.

Table 1237 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	804	967
HUMOSTRO_PEA_1_PEA_1_T16	727	890

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_31 according to the present invention is supported by 350 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1238 below describes the starting and ending position of this segment on each transcript.

Table 1238- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1164	1393

1250

HUMOSTRO_PEA_1_PEA_1_T16	1087	1316

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_43 according to the present invention is supported by 192 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1239 below describes the starting and ending position of this segment on each transcript.

Table 1239 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1810	1846
HUMOSTRO_PEA_1_PEA_1_T16	1733	1769

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_3 according to the present invention is supported by 353 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14, HUMOSTRO_PEA_1_PEA_1_T16 and HUMOSTRO_PEA_1_PEA_1_T30. Table 1240 below describes the starting and ending position of this segment on each transcript.

Table 1240- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	185	210
HUMOSTRO_PEA_1_PEA_1_T16	185	210
HUMOSTRO_PEA_1_PEA_1_T30	185	210

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_5 according to the present invention is supported by 353 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14, HUMOSTRO_PEA_1_PEA_1_T16 and HUMOSTRO_PEA_1_PEA_1_T30. Table 1241 below describes the starting and ending position of this segment on each transcript.

Table 1241- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	211	252
HUMOSTRO_PEA_1_PEA_1_T16	211	252
HUMOSTRO_PEA_1_PEA_1_T30	211	252

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_7 according to the present invention is supported by 357 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14, HUMOSTRO_PEA_1_PEA_1_T16 and HUMOSTRO_PEA_1_PEA_1_T30. Table 1242 below describes the starting and ending position of this segment on each transcript.

Table 1242- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	253	291
HUMOSTRO_PEA_1_PEA_1_T16	253	291
HUMOSTRO_PEA_1_PEA_1_T30	253	291

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_8 according to the present invention is supported by 1 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T30. Table 1243 below describes the starting and ending position of this segment on each transcript.

Table 1243- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T30	292	378

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_15 according to the present invention is supported by 366 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1244 below describes the starting and ending position of this segment on each transcript.

Table 1244 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	292	372
HUMOSTRO_PEA_1_PEA_1_T16	481	561

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_17 according to the present invention is supported by 261 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1245 below describes the starting and ending position of this segment on each transcript.

Table 1245 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	639	680
HUMOSTRO_PEA_1_PEA_1_T16	562	603

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_20 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1246 below describes the starting and ending position of this segment on each transcript.

Table 1246 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	681	688
HUMOSTRO_PEA_1_PEA_1_T16	604	611

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_21 according to the present invention is supported by 315 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1247 below describes the starting and ending position of this segment on each transcript.

Table 1247 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	689	738
HUMOSTRO_PEA_1_PEA_1_T16	612	661

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_22 according to the present invention is supported by 322 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1248 below describes the starting and ending position of this segment on each transcript.

Table 1248 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	739	803
HUMOSTRO_PEA_1_PEA_1_T16	662	726

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_24 according to the present invention is supported by 270 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1249 below describes the starting and ending position of this segment on each transcript.

Table 1249 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	968	1004
HUMOSTRO_PEA_1_PEA_1_T16	891	927

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_26 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1250 below describes the starting and ending position of this segment on each transcript.

Table 1250 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1005	1022
HUMOSTRO_PEA_1_PEA_1_T16	928	945

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_27 according to the present invention is supported by 260 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1251 below describes the starting and ending position of this segment on each transcript.

Table 1251 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1023	1048
HUMOSTRO_PEA_1_PEA_1_T16	946	971

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_28 according to the present invention is supported by 273 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1252 below describes the starting and ending position of this segment on each transcript.

Table 1252- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1049	1100
HUMOSTRO_PEA_1_PEA_1_T16	972	1023

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_29 according to the present invention is supported by 272 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1253 below describes the starting and ending position of this segment on each transcript.

Table 1253- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1101	1151
HUMOSTRO_PEA_1_PEA_1_T16	1024	1074

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_30 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1254 below describes the starting and ending position of this segment on each transcript.

Table 1254- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1152	1163
HUMOSTRO_PEA_1_PEA_1_T16	1075	1086

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_32 according to the present invention is supported by 293 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1255 below describes the starting and ending position of this segment on each transcript.

Table 1255- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1394	1427
HUMOSTRO_PEA_1_PEA_1_T16	1317	1350

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_34 according to the present invention is supported by 301 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1256 below describes the starting and ending position of this segment on each transcript.

Table 1256 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1428	1468
HUMOSTRO_PEA_1_PEA_1_T16	1351	1391

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_36 according to the present invention is supported by 292 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1257 below describes the starting and ending position of this segment on each transcript.

Table 1257 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1469	1504
HUMOSTRO_PEA_1_PEA_1_T16	1392	1427

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_37 according to the present invention is supported by 295 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1258 below describes the starting and ending position of this segment on each transcript.

Table 1258- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1505	1623
HUMOSTRO_PEA_1_PEA_1_T16	1428	1546

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_38 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1259 below describes the starting and ending position of this segment on each transcript.

Table 1259 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1624	1634
HUMOSTRO_PEA_1_PEA_1_T16	1547	1557

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_39 according to the present invention is supported by 268 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1260 below describes the starting and ending position of this segment on each transcript.

Table 1260 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1635	1725
HUMOSTRO_PEA_1_PEA_1_T16	1558	1648

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_40 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1261 below describes the starting and ending position of this segment on each transcript.

Table 1261 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1726	1743
HUMOSTRO_PEA_1_PEA_1_T16	1649	1666

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_41 according to the present invention can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1262 below describes the starting and ending position of this segment on each transcript.

Table 1262 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1744	1749
HUMOSTRO_PEA_1_PEA_1_T16	1667	1672

1260

Segment cluster HUMOSTRO_PEA_1_PEA_1_node_42 according to the present invention is supported by 224 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): HUMOSTRO_PEA_1_PEA_1_T14 and HUMOSTRO_PEA_1_PEA_1_T16. Table 1263 below describes the starting and ending position of this segment on each transcript.

Table 1263 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
HUMOSTRO_PEA_1_PEA_1_T14	1750	1809
HUMOSTRO_PEA_1_PEA_1_T16	1673	1732

Variant protein alignment to the previously known protein:

```
Sequence name: OSTP_HUMAN

Sequence documentation:

Alignment of: HUMOSTRO_PEA_1_PEA_1_P21 x OSTP_HUMAN    ..

Alignment segment 1/1:

                      Quality:  578.00
Escore:        0
              Matching length:       58                  Total
length:       58
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
```

1261

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQ 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQ 50

     51 KQNLLAPQ                                           58
        ||||||||
     51 KQNLLAPQ                                           58
```

Sequence name: OSTP_HUMAN

Sequence documentation:

Alignment of: HUMOSTRO_PEA_1_PEA_1_P25 x OSTP_HUMAN ..

Alignment segment 1/1:

Quality: 301.00

Escore: 0

Matching length: 31 Total length: 31

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .          .          .
        1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQ                    31
          ||||||||||||||||||||||||||||||||
        1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQ                    31
```

Sequence name: OSTP_HUMAN

Sequence documentation:

Alignment of: HUMOSTRO_PEA_1_PEA_1_P30 x OSTP_HUMAN ..

Alignment segment 1/1:

Quality: 301.00

Escore: 0

Matching length: 31 Total length: 31

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

1263

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .          .          .
        1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQ                31
          ||||||||||||||||||||||||||||||||
        1 MRIAVICFCLLGITCAIPVKQADSGSSEEKQ                31
```

DESCRIPTION FOR CLUSTER R11723

Cluster R11723 features 6 transcript(s) and 26 segment(s) of interest, the names for which are given in Tables 1264 and 1265, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1266.

Table 1264 - Transcripts of interest

Transcript Name	Sequence ID No.
R11723_PEA_1_T15	144
R11723_PEA_1_T17	145
R11723_PEA_1_T19	146
R11723_PEA_1_T20	147
R11723_PEA_1_T5	148
R11723_PEA_1_T6	149

Table 1265 - Segments of interest

Segment Name	Sequence ID No.
R11723_PEA_1_node_13	991
R11723_PEA_1_node_16	992
R11723_PEA_1_node_19	993
R11723_PEA_1_node_2	994
R11723_PEA_1_node_22	995
R11723_PEA_1_node_31	996
R11723_PEA_1_node_10	997
R11723_PEA_1_node_11	998
R11723_PEA_1_node_15	999
R11723_PEA_1_node_18	1000
R11723_PEA_1_node_20	1001
R11723_PEA_1_node_21	1002
R11723_PEA_1_node_23	1003
R11723_PEA_1_node_24	1004
R11723_PEA_1_node_25	1005
R11723_PEA_1_node_26	1006
R11723_PEA_1_node_27	1007
R11723_PEA_1_node_28	1008
R11723_PEA_1_node_29	1009
R11723_PEA_1_node_3	1010
R11723_PEA_1_node_30	1011
R11723_PEA_1_node_4	1012
R11723_PEA_1_node_5	1013
R11723_PEA_1_node_6	1014
R11723_PEA_1_node_7	1015
R11723_PEA_1_node_8	1016

Table 1266- Proteins of interest

Protein Name	Sequence ID No.
R11723_PEA_1_P2	1409
R11723_PEA_1_P6	1410
R11723_PEA_1_P7	1411
R11723_PEA_1_P13	1412
R11723_PEA_1_P10	1413

Cluster R11723 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 47 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 47 and Table 1267. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues and kidney malignant tumors.

Table 1267 - Normal tissue distribution

Name of Tissue	Number
Adrenal	0
Brain	30
Epithelial	3
General	17
head and neck	0
Kidney	0
Lung	0

Breast	0
Ovary	0
Pancreas	10
Skin	0
Uterus	0

Table 1268 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
Adrenal	4.2e-01	4.6e-01	4.6e-01	2.2	5.3e-01	1.9
Brain	2.2e-01	2.0e-01	1.2e-02	2.8	5.0e-02	2.0
Epithelial	3.0e-05	6.3e-05	1.8e-05	6.3	3.4e-06	6.4
General	7.2e-03	4.0e-02	1.3e-04	2.1	1.1e-03	1.7
head and neck	1	5.0e-01	1	1.0	7.5e-01	1.3
Kidney	1.5e-01	2.4e-01	4.4e-03	5.4	2.8e-02	3.6
Lung	1.2e-01	1.6e-01	1	1.6	1	1.3
Breast	5.9e-01	4.4e-01	1	1.1	6.8e-01	1.5
Ovary	1.6e-02	1.3e-02	1.0e-01	3.8	7.0e-02	3.5
Pancreas	5.5e-01	2.0e-01	3.9e-01	1.9	1.4e-01	2.7
Skin	1	4.4e-01	1	1.0	1.9e-02	2.1
Uterus	1.5e-02	5.4e-02	1.9e-01	3.1	1.4e-01	2.5

As noted above, contig R11723 features 6 transcript(s), which were listed in Table 1 above. A description of each variant protein according to the present invention is now provided.

Variant protein R11723_PEA_1_P2 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R11723_PEA_1_T6. The location of the variant protein was determined according to results

from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein R11723_PEA_1_P2 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1269, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1269 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
107	H -> P	Yes
70	G ->	No
70	G -> C	No

Variant protein R11723_PEA_1_P2 is encoded by the following transcript(s): R11723_PEA_1_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R11723_PEA_1_T6 is shown in bold; this coding portion starts at position 1716 and ends at position 2051. The transcript also has the following SNPs as listed in Table 1270 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P2 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1270 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
1231	C -> T	Yes
1278	G -> C	Yes
1923	G ->	No
1923	G -> T	No
2035	A -> C	Yes
2048	A -> C	No
2057	A -> G	Yes

Variant protein R11723_PEA_1_P6 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R11723_PEA_1_T15. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R11723_PEA_1_P6 and Q8IXM0 (SEQ ID NO:1707):

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence

MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV

MEQSAGIMYRKSCASSAACLIASAGSPCRGLAPGREEQRALHKAGAVGGGVR

corresponding to amino acids 1 - 110 of R11723_PEA_1_P6, and a second amino acid sequence being at least 90 % homologous to

MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLREGEEDHV

RPEVGPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRERQRKEKHSMRTQ

corresponding to amino acids 1 - 112 of Q8IXM0, which also corresponds to amino acids 111 -

222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAGSPCRGLAPGREEQRALHKAGAVGGGVR of R11723_PEA_1_P6.

Comparison report between R11723_PEA_1_P6 and Q96AC2 (SEQ ID NO:1708):

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 1 - 83 of Q96AC2, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

Comparison report between R11723_PEA_1_P6 and Q8N2G4 (SEQ ID NO:1709):

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 1 - 83 of Q8N2G4, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

Comparison report between R11723_PEA_1_P6 and BAC85518 (SEQ ID NO:1710):

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P6, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAGIMYRKSCASSAACLIASAG corresponding to amino acids 24 - 106 of BAC85518, which also corresponds to amino acids 1 - 83 of R11723_PEA_1_P6, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL

RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ corresponding to amino acids 84 - 222 of R11723_PEA_1_P6, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P6, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLL RGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQ CHNNQPWADTSRRERQRKEKHSMRTQ in R11723_PEA_1_P6.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein R11723_PEA_1_P6 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table , (given according to their 1271 position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1271- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
180	G ->	No
180	G -> C	No
217	H -> P	Yes

Variant protein R11723_PEA_1_P6 is encoded by the following transcript(s): R11723_PEA_1_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R11723_PEA_1_T15 is shown in bold; this coding portion starts at position 434 and ends at position 1099. The transcript also has the following SNPs as listed in Table 1272 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P6 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1272 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
971	G ->	No
971	G -> T	No
1083	A -> C	Yes
1096	A -> C	No
1105	A -> G	Yes

Variant protein R11723_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R11723_PEA_1_T17. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R11723_PEA_1_P7 and Q96AC2:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAG corresponding to amino acids 1 - 64 of Q96AC2, which also corresponds to amino

acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

Comparison report between R11723_PEA_1_P7 and Q8N2G4:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAG corresponding to amino acids 1 - 64 of Q8N2G4, which also corresponds to amino acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

Comparison report between R11723_PEA_1_P7 and BAC85273:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MWVLG corresponding to amino acids 1 - 5 of R11723_PEA_1_P7, second

amino acid sequence being at least 90 % homologous to IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSAG corresponding to amino acids 22 - 80 of BAC85273, which also corresponds to amino acids 6 - 64 of R11723_PEA_1_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLG of R11723_PEA_1_P7.

3.An isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

Comparison report between R11723_PEA_1_P7 and BAC85518:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSAG corresponding to amino acids 24 - 87 of BAC85518, which also corresponds to amino acids 1 - 64 of R11723_PEA_1_P7, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT corresponding to amino acids 65 - 93 of R11723_PEA_1_P7, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%,

more preferably at least about 90% and most preferably at least about 95% homologous to the sequence SHCVTRLECSGTISAHCNLCLPGSNDHPT in R11723_PEA_1_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein R11723_PEA_1_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1273, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1273- Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
67	C -> S	Yes

Variant protein R11723_PEA_1_P7 is encoded by the following transcript(s): R11723_PEA_1_T17, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R11723_PEA_1_T17 is shown in bold; this coding portion starts at position 434 and ends at position 712. The transcript also has the following SNPs as listed in Table 1274 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1274- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
625	G -> T	Yes
633	G -> C	Yes
1303	C -> T	Yes

Variant protein R11723_PEA_1_P13 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R11723_PEA_1_T19. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R11723_PEA_1_P13 and Q96AC2:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P13, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q96AC2, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P13, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DTKRTNTLLFEMRHFAKQLTT corresponding to amino acids 64 - 84 of R11723_PEA_1_P13, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P13, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DTKRTNTLLFEMRHFAKQLTT in R11723_PEA_1_P13.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell:

secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein R11723_PEA_1_P13 is encoded by the following transcript(s): R11723_PEA_1_T19 and R11723_PEA_1_T5, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R11723_PEA_1_T19 is shown in bold; this coding portion starts at position 434 and ends at position 685. The transcript also has the following SNPs as listed in Table 1275 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P13 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1275 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
778	G -> T	Yes
786	G -> C	Yes
1456	C -> T	Yes

Variant protein R11723_PEA_1_P10 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R11723_PEA_1_T20. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R11723_PEA_1_P10 and Q96AC2:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to

MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q96AC2, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

Comparison report between R11723_PEA_1_P10 and Q8N2G4:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q8N2G4, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

Comparison report between R11723_PEA_1_P10 and BAC85273:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence MWVLG corresponding to amino acids 1 - 5 of R11723_PEA_1_P10, second amino acid sequence being at least 90 % homologous to IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSA corresponding to amino acids 22 - 79 of BAC85273, which also corresponds to amino acids 6 - 63 of R11723_PEA_1_P10, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of R11723_PEA_1_P10, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a head of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence MWVLG of R11723_PEA_1_P10.

3.An isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

Comparison report between R11723_PEA_1_P10 and BAC85518:

1.An isolated chimeric polypeptide encoding for R11723_PEA_1_P10, comprising a first amino acid sequence being at least 90 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 24 - 86 of BAC85518, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P10, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK corresponding to amino acids 64 - 90 of

R11723_PEA_1_P10, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R11723_PEA_1_P10, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence DRVSLCHEAGVQWNNFSTLQPLPPRLK in R11723_PEA_1_P10.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein R11723_PEA_1_P10 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1276, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R11723_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1276 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
66	V -> F	Yes

Variant protein R11723_PEA_1_P10 is encoded by the following transcript(s): R11723_PEA_1_T20, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R11723_PEA_1_T20 is shown in bold; this coding portion starts at position 434 and ends at position 703. The transcript also has the following SNPs as listed in Table 1277 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein R11723_PEA_1_P10 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1277- Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
629	G -> T	Yes
637	G -> C	Yes
1307	C -> T	Yes

As noted above, cluster R11723 features 26 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R11723_PEA_1_node_13 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T19, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1278 below describes the starting and ending position of this segment on each transcript.

Table 1278- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T19	624	776
R11723_PEA_1_T5	624	776
R11723_PEA_1_T6	658	810

Segment cluster R11723_PEA_1_node_16 according to the present invention is supported by 3 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T17, R11723_PEA_1_T19 and

R11723_PEA_1_T20. Table 1279 below describes the starting and ending position of this segment on each transcript.

Table 1279- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T17	624	1367
R11723_PEA_1_T19	777	1520
R11723_PEA_1_T20	628	1371

Segment cluster R11723_PEA_1_node_19 according to the present invention is supported by 45 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1280 below describes the starting and ending position of this segment on each transcript.

Table 1280- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T5	835	1008
R11723_PEA_1_T6	869	1042

Segment cluster R11723_PEA_1_node_2 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1281 below describes the starting and ending position of this segment on each transcript.

Table 1281- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	1	309
R11723_PEA_1_T17	1	309
R11723_PEA_1_T19	1	309

R11723_PEA_1_T20	1	309
R11723_PEA_1_T5	1	309
R11723_PEA_1_T6	1	309

Segment cluster R11723_PEA_1_node_22 according to the present invention is supported by 65 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1282 below describes the starting and ending position of this segment on each transcript.

Table 1282- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T5	1083	1569
R11723_PEA_1_T6	1117	1603

Segment cluster R11723_PEA_1_node_31 according to the present invention is supported by 70 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1283 below describes the starting and ending position of this segment on each transcript (it should be noted that these transcripts show alternative polyadenylation).

Table 1283 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	1060	1295
R11723_PEA_1_T5	1978	2213
R11723_PEA_1_T6	2012	2247

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R11723_PEA_1_node_10 according to the present invention is supported by 38 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1284 below describes the starting and ending position of this segment on each transcript.

Table 1284 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	486	529
R11723_PEA_1_T17	486	529
R11723_PEA_1_T19	486	529
R11723_PEA_1_T20	486	529
R11723_PEA_1_T5	486	529
R11723_PEA_1_T6	520	563

Segment cluster R11723_PEA_1_node_11 according to the present invention is supported by 42 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1285 below describes the starting and ending position of this segment on each transcript.

Table 1285 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	530	623
R11723_PEA_1_T17	530	623
R11723_PEA_1_T19	530	623
R11723_PEA_1_T20	530	623
R11723_PEA_1_T5	530	623
R11723_PEA_1_T6	564	657

Segment cluster R11723_PEA_1_node_15 according to the present invention can be found in the following transcript(s): R11723_PEA_1_T20. Table 1286 below describes the starting and ending position of this segment on each transcript.

Table 1286 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T20	624	627

Segment cluster R11723_PEA_1_node_18 according to the present invention is supported by 40 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1287 below describes the starting and ending position of this segment on each transcript.

Table 1287- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	624	681
R11723_PEA_1_T5	777	834
R11723_PEA_1_T6	811	868

Segment cluster R11723_PEA_1_node_20 according to the present invention can be found in the following transcript(s): R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1288 below describes the starting and ending position of this segment on each transcript.

Table 1288- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T5	1009	1019
R11723_PEA_1_T6	1043	1053

Segment cluster R11723_PEA_1_node_21 according to the present invention is supported by 36 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1289 below describes the starting and ending position of this segment on each transcript.

Table 1289 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T5	1020	1082
R11723_PEA_1_T6	1054	1116

Segment cluster R11723_PEA_1_node_23 according to the present invention is supported by 39 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1290 below describes the starting and ending position of this segment on each transcript.

Table 1290 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T5	1570	1599
R11723_PEA_1_T6	1604	1633

Segment cluster R11723_PEA_1_node_24 according to the present invention is supported by 51 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1291 below describes the starting and ending position of this segment on each transcript.

Table 1291 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	682	765

R11723_PEA_1_T5	1600	1683
R11723_PEA_1_T6	1634	1717

Segment cluster R11723_PEA_1_node_25 according to the present invention is supported by 54 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1292 below describes the starting and ending position of this segment on each transcript.

Table 1292 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	766	791
R11723_PEA_1_T5	1684	1709
R11723_PEA_1_T6	1718	1743

Segment cluster R11723_PEA_1_node_26 according to the present invention is supported by 62 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1293 below describes the starting and ending position of this segment on each transcript.

Table 1293 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	792	904
R11723_PEA_1_T5	1710	1822
R11723_PEA_1_T6	1744	1856

Segment cluster R11723_PEA_1_node_27 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1294 below describes the starting and ending position of this segment on each transcript.

Table 1294 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	905	986
R11723_PEA_1_T5	1823	1904
R11723_PEA_1_T6	1857	1938

Segment cluster R11723_PEA_1_node_28 according to the present invention can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1295 below describes the starting and ending position of this segment on each transcript.

Table 1295 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	987	1010
R11723_PEA_1_T5	1905	1928
R11723_PEA_1_T6	1939	1962

Segment cluster R11723_PEA_1_node_29 according to the present invention is supported by 69 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1296 below describes the starting and ending position of this segment on each transcript.

Table 1296 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	1011	1038

R11723_PEA_1_T5	1929	1956
R11723_PEA_1_T6	1963	1990

Segment cluster R11723_PEA_1_node_3 according to the present invention can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1297 below describes the starting and ending position of this segment on each transcript.

Table 1297 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	310	319
R11723_PEA_1_T17	310	319
R11723_PEA_1_T19	310	319
R11723_PEA_1_T20	310	319
R11723_PEA_1_T5	310	319
R11723_PEA_1_T6	310	319

Segment cluster R11723_PEA_1_node_30 according to the present invention can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1298 below describes the starting and ending position of this segment on each transcript.

Table 1298 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	1039	1059
R11723_PEA_1_T5	1957	1977
R11723_PEA_1_T6	1991	2011

Segment cluster R11723_PEA_1_node_4 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1299 below describes the starting and ending position of this segment on each transcript.

Table 1299 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	320	371
R11723_PEA_1_T17	320	371
R11723_PEA_1_T19	320	371
R11723_PEA_1_T20	320	371
R11723_PEA_1_T5	320	371
R11723_PEA_1_T6	320	371

Segment cluster R11723_PEA_1_node_5 according to the present invention is supported by 26 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1300 below describes the starting and ending position of this segment on each transcript.

Table 1300 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	372	414
R11723_PEA_1_T17	372	414
R11723_PEA_1_T19	372	414
R11723_PEA_1_T20	372	414
R11723_PEA_1_T5	372	414
R11723_PEA_1_T6	372	414

Segment cluster R11723_PEA_1_node_6 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1301 below describes the starting and ending position of this segment on each transcript.

Table 1301 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	415	446
R11723_PEA_1_T17	415	446
R11723_PEA_1_T19	415	446
R11723_PEA_1_T20	415	446
R11723_PEA_1_T5	415	446
R11723_PEA_1_T6	415	446

Segment cluster R11723_PEA_1_node_7 according to the present invention is supported by 29 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T15, R11723_PEA_1_T17, R11723_PEA_1_T19, R11723_PEA_1_T20, R11723_PEA_1_T5 and R11723_PEA_1_T6. Table 1302 below describes the starting and ending position of this segment on each transcript.

Table 1302 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T15	447	485
R11723_PEA_1_T17	447	485
R11723_PEA_1_T19	447	485
R11723_PEA_1_T20	447	485
R11723_PEA_1_T5	447	485
R11723_PEA_1_T6	447	485

Segment cluster R11723_PEA_1_node_8 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R11723_PEA_1_T6. Table 1303 below describes the starting and ending position of this segment on each transcript.

Table 1303 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R11723_PEA_1_T6	486	519

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/gp6eQTLWqk/mFtjUpUzhb:Q8IXM0

Sequence documentation:

Alignment of: R11723_PEA_1_P6 x Q8IXM0    ..

Alignment segment 1/1:

                        Quality: 1128.00
Escore:        0
               Matching length:     112                Total
length:      112
```

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
     111 MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLRE 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLRE 50
                .         .         .         .         .
     161 GEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRE 210
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRE 100
                .
     211 RQRKEKHSMRTQ                                      222
         ||||||||||||
     101 RQRKEKHSMRTQ                                      112
```

Sequence name: /tmp/gp6eQTLWqk/mFtjUpUzhb:Q96AC2

Sequence documentation:

Alignment of: R11723_PEA_1_P6 x Q96AC2 ..

```
Alignment segment 1/1:

                    Quality:  835.00
Escore:       0
              Matching length:      83                Total
length:      83
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                        Gaps:       0

Alignment:
                 .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
                 .         .         .
       51 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  83
          |||||||||||||||||||||||||||||||||
       51 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  83


Sequence name: /tmp/gp6eQTLWqk/mFtjUpUzhb:Q8N2G4

Sequence documentation:
```

```
Alignment of: R11723_PEA_1_P6 x Q8N2G4    ..

Alignment segment 1/1:

                      Quality:  835.00
Escore:       0
             Matching length:      83                 Total
length:      83
 Matching Percent Similarity:  100.00   Matching Percent
Identity:  100.00
    Total Percent Similarity:  100.00      Total Percent
Identity:  100.00
                         Gaps:       0

Alignment:
                  .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
                  .         .         .
       51 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  83
          |||||||||||||||||||||||||||||||||
       51 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  83


Sequence name: /tmp/gp6eQTLWqk/mFtjUpUzhb:BAC85518
```

Sequence documentation:

Alignment of: R11723_PEA_1_P6 x BAC85518 ..

Alignment segment 1/1:

Quality: 835.00

Escore: 0

Matching length: 83 Total length: 83

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       24 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 73
                .         .         .
       51 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  83
          |||||||||||||||||||||||||||||||||
       74 QDMCQKEVMEQSAGIMYRKSCASSAACLIASAG                  106
```

Sequence name: /tmp/VXjdFlzdBX/bexTxTh0Th:Q96AC2

Sequence documentation:

Alignment of: R11723_PEA_1_P7 x Q96AC2 ..

Alignment segment 1/1:

Quality: 654.00
Escore: 0
Matching length: 64 Total length: 64
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
                  .
       51 QDMCQKEVMEQSAG                                    64
          ||||||||||||||
       51 QDMCQKEVMEQSAG                                    64
```

Sequence name: /tmp/VXjdFlzdBX/bexTxTh0Th:Q8N2G4

Sequence documentation:

Alignment of: R11723_PEA_1_P7 x Q8N2G4 ..

Alignment segment 1/1:

Quality: 654.00

Escore: 0

Matching length: 64 Total length: 64

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
                .
     51 QDMCQKEVMEQSAG                                    64
        ||||||||||||||
     51 QDMCQKEVMEQSAG                                    64
```

Sequence name: /tmp/VXjdFlzdBX/bexTxTh0Th:BAC85273

Sequence documentation:

Alignment of: R11723_PEA_1_P7 x BAC85273 ..

Alignment segment 1/1:

Quality: 600.00
Escore: 0
Matching length: 59 Total length: 59
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                    .         .         .         .         .
         6 IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQ 55
           |||||||||||||||||||||||||||||||||||||||||||||||||||
        22 IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQ 71

        56 KEVMEQSAG                                          64
           |||||||||
        72 KEVMEQSAG                                          80
```

1300

Sequence name: /tmp/VXjdFlzdBX/bexTxTh0Th:BAC85518

Sequence documentation:

Alignment of: R11723_PEA_1_P7 x BAC85518 ..

Alignment segment 1/1:

Quality: 654.00
Escore: 0
Matching length: 64 Total length: 64
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                 .         .         .         .         .
       1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      24 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 73
                 .
      51 QDMCQKEVMEQSAG                                    64
```

1301

```
            ||||||||||||||
        74 QDMCQKEVMEQSAG                                        87
```

Sequence name: /tmp/OLMSexEmIh/pc7Z7Xm1YR:Q96AC2

Sequence documentation:

Alignment of: R11723_PEA_1_P10 x Q96AC2 ..

Alignment segment 1/1:

Quality: 645.00

Escore: 0

Matching length: 63 Total length: 63

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
```

```
        51 QDMCQKEVMEQSA                                                63
           |||||||||||||
        51 QDMCQKEVMEQSA                                                63
```

Sequence name: /tmp/OLMSexEmIh/pc7Z7Xm1YR:Q8N2G4

Sequence documentation:

Alignment of: R11723_PEA_1_P10 x Q8N2G4 ..

Alignment segment 1/1:

Quality: 645.00
Escore: 0
Matching length: 63 Total length: 63
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
         1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
```

1303

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
          1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
                          .
         51 QDMCQKEVMEQSA                                      63
5           ||||||||||||
         51 QDMCQKEVMEQSA                                      63
```

10

Sequence name: /tmp/OLMSexEmIh/pc7Z7Xm1YR:BAC85273

15 Sequence documentation:

Alignment of: R11723_PEA_1_P10 x BAC85273 ..

Alignment segment 1/1:

20

Quality: 591.00
Escore: 0
Matching length: 58 Total length: 58
25 Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

30

Alignment:

```
                .         .         .         .         .
 6 IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQ 55
   ||||||||||||||||||||||||||||||||||||||||||||||||||
22 IAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQ 71

56 KEVMEQSA                                           63
   ||||||||
72 KEVMEQSA                                           79
```

Sequence name: /tmp/OLMSexEmIh/pc7Z7Xm1YR:BAC85518

Sequence documentation:

Alignment of: R11723_PEA_1_P10 x BAC85518 ..

Alignment segment 1/1:

Quality: 645.00
Escore: 0
Matching length: 63 Total length: 63
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

```
Alignment:
                    .         .         .         .         .
        1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       24 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 73
                   .
       51 QDMCQKEVMEQSA                                      63
          |||||||||||||
       74 QDMCQKEVMEQSA                                      86
```

Alignment of: R11723_PEA_1_P13 x Q96AC2 ..

Alignment segment 1/1:

Quality: 645.00

Escore: 0

Matching length: 63 Total length: 63

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNV 50

                .
 51 QDMCQKEVMEQSA                                      63
    |||||||||||||
 51 QDMCQKEVMEQSA                                      63
```

It should be noted that the nucleotide transcript sequence of known protein (PSEC, also referred to herein as the "wild type" or WT protein) feature at least one SNP that appears to affect the coding region, in addition to certain silent SNPs. This SNP does not have an effect on the R11723_PEA_1_T5 splice variant sequence): "G-> " resulting in a missing nucleotide (affects amino acids from position 91 onwards). The missing nucleotide creates a frame shift, resulting in a new protein. This SNP was not previously identified and is supported by 5 ESTs out of ~70 ESTs in this exon.

It should be noted that the variants of this cluster are variants of the hypothetical protein PSEC0181 (referred to herein as "PSEC"). Furthermore, use of the known protein (WT protein) for detection of lung cancer, alone or in combination with one or more variants of this cluster and/or of any other cluster and/or of any known marker, also comprises an embodiment of the present invention.

Expression of R11723 transcripts which are detectable by amplicon as depicted in sequence name R11723 seg13 in normal and cancerous lung tissues

Expression of transcripts detectable by or according to R11723 seg13, R11723 seg13 amplicon (SEQ ID NO: 1684), and R11723 seg13F (SEQ ID NO: 1682), and R11723 seg13R (SEQ ID NO: 1683), primers was measured by real time PCR. In parallel the expression of four housekeeping genes PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), and SDHA (GenBank Accession No. NM_004168; amplicon –

SDHA-amplicon, SEQ ID NO:331), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2 "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 48 is a histogram showing over expression of the above-indicated transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 5 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 48, the expression of transcripts detectable by the above amplicon(s) in cancer samples was higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2 "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 10 out of 15 adenocarcinoma samples, and in 4 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: R11723 seg13F forward primer; and R11723 seg13R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: R11723 seg13.

R11723seg13F (SEQ ID NO: 1682), - ACACTAAAAGAACAAACACCTTGCTC

R11723seg13R (SEQ ID NO: 1683), - TCCTCAGAAGGCACATGAAAGA

R11723seg13 – amplicon (SEQ ID NO: 1684),:

ACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCA GTTGACCACTTAGTTCTCAAGAAGCAACTATCTCTTTCATGTGCCTTCTGAGGA

Expression of R11723 transcripts which are detectable by amplicon as depicted in sequence name R11723seg13 in different normal tissues

Expression of R11723 transcripts detectable by or according to R11723seg13 amplicon (SEQ ID NO: 1684), and R11723seg13F (SEQ ID NO: 1682),, R11723seg13R (SEQ ID NO: 1683), was measured by real time PCR. In parallel the expression of four housekeeping genes RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20, Table 2 "Tissue samples in normal panel" above), to obtain a value of relative expression of each sample relative to median of the ovary samples.

R11723seg13F (SEQ ID NO: 1682), - ACACTAAAAGAACAAACACCTTGCTC

R11723seg13R (SEQ ID NO: 1683), - TCCTCAGAAGGCACATGAAAGA

R11723seg13 – amplicon (SEQ ID NO: 1684),:

ACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCAGTTG

ACCACTTAGTTCTCAAGAAGCAACTATCTCTTTCATGTGCCTTCTGAGGA

The results are presented in Figure 49, showing the expression of R11723 transcripts which are detectable by amplicon as depicted in sequence name R11723seg13 in different normal tissues.

Expression of R11723 transcripts, which are detectable by amplicon as depicted in sequence name R11723 junc11-18 in normal and cancerous lung tissues.

Expression of transcripts detectable by or according to junc11-18, R11723 junc11-18 amplicon (SEQ ID NO: 1687) and R11723 junc11-18F (SEQ ID NO: 1685) and R11723 junc11-18R (SEQ ID NO: 1686) primers was measured by real time PCR (this junction is found in the known protein sequence or "wild type" (WT) sequence, also termed herein the PSEC sequence). In parallel the expression of four housekeeping genes PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No.

NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), and Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above: "Tissue samples in lung cancer testing panel"), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 50 is a histogram showing over expression of the above-indicated transcripts in cancerous lung samples relative to the normal samples. Values represent the average of duplicate experiments. Error bars indicate the minimal and maximal values obtained.

As is evident from Figure 50, the expression of transcripts detectable by the above amplicon in cancer samples was higher than in the non-cancerous samples (Sample Nos. 47-50, 90-93, 96-99 Table 2 "Tissue samples in lung cancer testing panel"). Notably an over-expression of at least 5 fold was found in 11 out of 15 adenocarcinoma samples, 4 out of 16 squamous cell carcinoma samples, 1 out of 4 large cell carcinoma samples and in 5 out of 8 small cells carcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: R11723 junc11-18F forward primer; and R11723 junc11-18R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: R11723 junc11-18.

R11723junc11-18F (SEQ ID NO: 1685)– AGTGATGGAGCAAAGTGCCG

R11723 junc11-18R (SEQ ID NO: 1686)- CAGCAGCTGATGCAAACTGAG

R11723 junc11-18 – amplicon (SEQ ID NO: 1687)

1310

AGTGATGGAGCAAAGTGCCGGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGG
CCTGTCTCATCGCCTCTGCCGGGTACCAGTCCTTCTGCTCCCCAGGGAAACTGAACT
CAGTTTGCATCAGCTGCTG

Expression of R11723 transcripts, which were detected by amplicon as depicted in the sequence name R11723 junc11-18 in different normal tissues.

Expression of *R11723* transcripts detectable by or according to *R11723seg13* amplicon (SEQ ID NO: 1687) and *R11723 junc11-18F* (SEQ ID NO: 1685), *R11723 junc11-18R*(SEQ ID NO: 1686) was measured by real time PCR. In parallel the expression of four housekeeping genes RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (Sample Nos. 18-20 Table 3 above), to obtain a value of relative expression of each sample relative to median of the ovary samples.

R11723junc11-18F (SEQ ID NO: 1685)– AGTGATGGAGCAAAGTGCCG
R11723 junc11-18R (SEQ ID NO: 1686)- CAGCAGCTGATGCAAACTGAG
R11723 junc11-18 – amplicon (SEQ ID NO: 1687)

AGTGATGGAGCAAAGTGCCGGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGG
CCTGTCTCATCGCCTCTGCCGGGTACCAGTCCTTCTGCTCCCCAGGGAAACTGAACT
CAGTTTGCATCAGCTGCTG

The results are demonstrated in Figure 73, showing the expression of R11723 transcripts, which were detected by amplicon as depicted in the sequence name R11723 junc11-18 in different normal tissues.

Cloning of this variant

Full length validation

RNA preparation

Human adult papillary adenocarcinoma ovary RNA pool (lot# ILS1408) was obtained from ABS (http://www.absbioreagents, Wilmington, DE 19801, USA com). Total RNA samples were treated with DNaseI (Ambion Cat # 1906).

RT PCR

RT preparation

Purified RNA (1 ug) was mixed with 150 ng Random Hexamer primers (Invitrogen Cat # 48190-011) and 500 uM dNTP (Takara, Cat # B9501-1) in a total volume of 15.6ul DEPC-H_2O (Beit Haemek, Cat # 01-852-1A). The mixture was incubated for 5 min at 65°C and then quickly chilled on ice. Thereafter, 5 ul of 5X Superscript II first strand buffer (Invitrogen, Cat # Y00146), 2.4ul 0.1M DTT (Invitrogen, Cat #Y00147) and 40 units RNasin (Promega, Cat # N251A) were added, and the mixture was incubated for 2 min at 42°C. Then, 1 ul (200units) of SuperscriptII (Invitrogen, Cat #18064-022) was added and the reaction was incubated for 50 min at 42°C and then inactivated at 70°C for 15min. The resulting cDNA was diluted 1:20 in TE buffer (10 mM Tris pH=8, 1 mM EDTA pH=8).

PCR amplification and analysis

cDNA (5ul), prepared as described above, was used as a template in PCR reactions. The amplification was done using AccuPower PCR PreMix (Bioneer, Korea, Cat# K2016), under the following conditions: 1ul – of each primer (10uM)

PSECfor- TGCTGTCGCCTCCTCTGATG

PSECrev- CCTCAGAAGGCACATGAAAG

plus 13ul – H_2O were added into AccuPower PCR PreMix tube with a reaction program of 5 minutes at 94°C; 35 cycles of: [30 seconds at 94°C, 30 seconds at 52°C, 40 seconds at 72°C] and 10 minutes at 72°C. At the end of the PCR amplification, products were analyzed on agarose gels stained with ethidium bromide and visualized with UV light. PCR product was extracted

from the gel using QiaQuick™ gel extraction kit (Qiagen™, Cat #28706). The extracted DNA product (Figure 79) was sequenced by direct sequencing using the gene specific primers from above (Hy-Labs, Israel), resulting in the expected sequence of PSEC variant R11723_PEA_1 T5 (Figure 80).

It was concluded that the predicted PSEC variant R11723_PEA_1 T5 is indeed a naturally expressed variant in an adult papillary adenocarcinoma ovary human tissue as shown in Figure 79.

Cloning of PSEC variant R11723_PEA_1 T5 into bacterial expression vector

The PSEC splice variant R11723_PEA_1 T5 coding sequence was prepared for cloning by PCR amplification using the fragment described above as template and Platinum Pfx DNA polymerase (Invitrogen Cat # 11708021) under the following conditions: 5?ul – Amplification X10 buffer (Invitrogen Cat # 11708021); 2ul – PCR product from above; 1ul – dNTPs (10mM each); 1µl MgSO4 (50mM) 5ul enhancer solution (Invitrogen Cat # 11708021); 33?ul – H_2O; 1ul – of each primer (10uM) and 1.25 units of *Taq* polymerase [Platinum Pfx DNA polymerase (Invitrogen Cat # 11708021)] in a total reaction volume of 50ul with a reaction program of 3 minutes at 94°C; 29 cycles of: [30 seconds at 94°C, 30 seconds at 58°C, 40 seconds at 68°C] and 7 minutes at 68 °C. The Primers listed below include specific sequences of the nucleotide sequence corresponding to the splice variant and NheI and HindIII restriction sites.

PSEC NheIfor- ATAGCTAGCATGTGGGTCCTAGGCATCGCGG

PSEC HindIIIrev- CCCAAGCTTCTAAGTGGTCAACTGCTTGGC

The PCR product was then double digested with NheI and HindIII (New England Biolabs (UK) LTD) (Figure 81), and inserted into pRSET-A (Invitrogen, Cat# V351-20), previously digested with the same enzymes, in-frame to an N-terminal 6His-tag, to give HisPSEC T5 pRSET (Figure 82). The coding sequence encodes for a protein having the 6His-tag at the N' end (6His residues in a row at one end of the protein), and 8 additional amino acids encoded by the pRSET vector.

The sequence of the PSEC insert in the final plasmid, as well as its flanking regions, were verified by sequencing and found to be identical to the desired sequences. The complete sequence of His PSEC T5 pRESTA, including the sequenced regions, is shown in Figure 84.

Figure 83 shows the translated sequence of PSEC variant R11723_PEA_1 T5.

Bacterial culture and induction of protein expression

HisPSEC pRSETA DNA was transformed into competent DH5a cells (Invitrogen Cat#18258-012). Ampicillin resistant transformants were screened and positive clones were further analyzed by restriction enzyme digestion and sequence verification.

In order to express the recombinant protein, HisPSEC pRSETA DNA was further transformed into competent BL21Gold cells (Stratagene Cat#230134) and BL21star (Invitrogen Cat# 44-0054). Ampicillin resistant transformants were screened and positive clones were selected.

Bacterial cells containing the HisPSEC T5 pRSET vector or empty pRSET vector (as negative control) were grown in LB medium, supplemented with Ampicillin (50 ug/ml) and chloramphenicol (34 ug/ml), until O.D.600nm reached 0.55. This value was reached in about 3 hours. 1mM IPTG (Roche, Cat #724815) was added and the cells were grown at 37°C overnight. 1 ml aliquots of each culture were removed for gel analysis at time zero, 3 hrs after induction and following overnight incubation (T0 ,T3 and TO/N, respectively).

Expression Results

The time course of small-scale expression of PSEC in BL21Gold is demonstrated in Figure 85. The expression of a recombinant protein with the appropriate molecular weight (9.2 kDa) was visualized by Western Blot with anti-His antibodies (BD Clontech, Ref 631212, Figure 85), but not by Coomassie staining (data not shown). Similar expression pattern was obtained with BL21 star as well (data not shown).

These results show that the protein encoded by PSEC variant R11723_PEA_1 T5 is indeed expressed in bacterial cells.

DESCRIPTION FOR CLUSTER R16276

Cluster R16276 features 1 transcript(s) and 5 segment(s) of interest, the names for which are given in Tables 1305 and 1306, respectively, the sequences themselves are given at the end of the application. The selected protein variants are given in table 1307.

Table 1305 - Transcripts of interest

Transcript Name	Sequence ID No.
R16276_PEA_1_T6	150

Table 1306 - Segments of interest

Segment Name	Sequence ID No.
R16276_PEA_1_node_0	1017
R16276_PEA_1_node_6	1018
R16276_PEA_1_node_1	1019
R16276_PEA_1_node_4	1020
R16276_PEA_1_node_5	1021

Table 1307 - Proteins of interest

Protein Name	Sequence ID No.	Corresponding Transcript(s)
R16276_PEA_1_P7	1414	R16276_PEA_1_T6

These sequences are variants of the known protein NOV protein homolog precursor (SwissProt accession identifier NOV_HUMAN; known also according to the synonyms NovH; Nephroblastoma overexpressed gene protein homolog), SEQ ID NO:1463, referred to herein as the previously known protein.

Protein NOV protein homolog precursor is known or believed to have the following function(s): Immediate-early protein, likely to play a role in cell growth regulation (By similarity). The sequence for protein NOV protein homolog precursor is given at the end of the application, as "NOV protein homolog precursor amino acid sequence". Known polymorphisms for this sequence are as shown in Table 1308.

Table 1308 - Amino acid mutations for Known Protein

SNP position(s) on amino acid sequence	Comment
97	N -> K

Protein NOV protein homolog precursor localization is believed to be Secreted.

The following GO Annotation(s) apply to the previously known protein. The following annotation(s) were found: regulation of cell growth, which are annotation(s) related to Biological Process; insulin-like growth factor binding; growth factor, which are annotation(s) related to Molecular Function; and extracellular, which are annotation(s) related to Cellular Component.

The GO assignment relies on information from one or more of the SwissProt/TremBl Protein knowledgebase, available from <http://www.expasy.ch/sprot/>; or Locuslink, available from <http://www.ncbi.nlm.nih.gov/projects/LocusLink/>.

Cluster R16276 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 51 refer to weighted expression of ESTs in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 51 and Table 1309. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: lung malignant tumors.

Table 1310 - Normal tissue distribution

Name of Tissue	Number
Adrenal	977
Bone	32
Brain	24

Colon	0
Epithelial	63
General	43
Kidney	24
Liver	341
Lung	0
Breast	0
Muscle	20
Ovary	0
Pancreas	0
Prostate	24
Skin	13
Stomach	146
Uterus	0

Table 1311 - P values and ratios for expression in cancerous tissue

Name of Tissue	P1	P2	SP1	R3	SP2	R4
Adrenal	5.9e-01	6.2e-01	1	0.2	9.9e-01	0.2
Bone	5.5e-01	7.3e-01	1	0.8	1	0.6
Brain	2.8e-01	4.4e-01	6.8e-01	0.9	8.9e-01	0.6
Colon	2.6e-01	3.3e-01	4.9e-01	2.0	5.9e-01	1.7
Epithelial	2.6e-01	2.9e-01	9.7e-01	0.6	1	0.5
General	4.1e-01	6.8e-01	9.4e-01	0.7	1	0.5
Kidney	8.3e-01	7.7e-01	6.2e-01	1.2	5.3e-01	1.4
Liver	9.1e-01	7.5e-01	1	0.1	1	0.1
Lung	2.3e-02	9.1e-02	8.0e-04	10.5	2.1e-02	5.1
Breast	5.9e-01	6.7e-01	6.9e-01	1.5	8.2e-01	1.2
Muscle	5.2e-01	6.1e-01	2.7e-01	3.2	6.3e-01	1.2
Ovary	6.2e-01	6.5e-01	6.8e-01	1.5	7.7e-01	1.3

Pancreas	3.3e-01	4.4e-01	4.2e-01	2.4	5.3e-01	1.9
Prostate	9.3e-01	9.4e-01	1	0.5	9.4e-01	0.6
Skin	9.2e-01	6.8e-01	1	0.5	4.1e-01	1.1
Stomach	5.0e-01	7.3e-01	5.0e-01	0.6	9.7e-01	0.4
Uterus	2.4e-01	1.6e-01	2.9e-01	2.5	4.1e-01	2.0

As noted above, cluster R16276 features 1 transcript(s), which were listed in Table 1 above. These transcript(s) encode for protein(s) which are variant(s) of protein NOV protein homolog precursor. A description of each variant protein according to the present invention is now provided.

Variant protein R16276_PEA_1_P7 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) R16276_PEA_1_T6. An alignment is given to the known protein (NOV protein homolog precursor) at the end of the application. One or more alignments to one or more previously published protein sequences are given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to each such aligned protein is as follows:

Comparison report between R16276_PEA_1_P7 and NOV_HUMAN:

1.An isolated chimeric polypeptide encoding for R16276_PEA_1_P7, comprising a first amino acid sequence being at least 90 % homologous to MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPG corresponding to amino acids 1 - 41 of NOV_HUMAN, which also corresponds to amino acids 1 - 41 of R16276_PEA_1_P7, a bridging amino acid Q corresponding to amino acid 42 of R16276_PEA_1_P7, a second amino acid sequence being at least 90 % homologous to CPATPPTCAPGVRAVLDGCSCCLVCARQRGESCSDLEPCDESSGLYCDRSADPSNQTGI CT corresponding to amino acids 43 - 103 of NOV_HUMAN, which also corresponds to amino acids 43 - 103 of R16276_PEA_1_P7, and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GNPAPSAV corresponding to amino acids 104 - 111 of R16276_PEA_1_P7, wherein said first amino acid

sequence, bridging amino acid, second amino acid sequence and third amino acid sequence are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of R16276_PEA_1_P7, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence GNPAPSAV in R16276_PEA_1_P7.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region.

Variant protein R16276_PEA_1_P7 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table 1312, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R16276_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1313 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
42	Q -> R	Yes

The glycosylation sites of variant protein R16276_PEA_1_P7, as compared to the known protein NOV protein homolog precursor, are described in Table 1314 (given according to their position(s) on the amino acid sequence in the first column; the second column indicates whether the glycosylation site is present in the variant protein; and the last column indicates whether the position is different on the variant protein).

Table 1314 - Glycosylation site(s)

Position(s) on known amino acid sequence	Present in variant protein?	Position in variant protein?
280	no	
97	yes	97

Variant protein R16276_PEA_1_P7 is encoded by the following transcript(s): R16276_PEA_1_T6, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript R16276_PEA_1_T6 is shown in bold; this coding portion starts at position 445 and ends at position 777. The transcript also has the following SNPs as listed in Table 1315 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein R16276_PEA_1_P7 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1315 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
371	G ->	No
430	A -> G	No
569	A -> G	Yes
729	C -> A	Yes
827	G -> T	Yes

As noted above, cluster R16276 features 5 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster R16276_PEA_1_node_0 according to the present invention is supported by 35 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): R16276_PEA_1_T6. Table 1316 below describes the starting and ending position of this segment on each transcript.

Table 1316 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R16276_PEA_1_T6	1	438

Segment cluster R16276_PEA_1_node_6 according to the present invention is supported by 2 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R16276_PEA_1_T6. Table 1317 below describes the starting and ending position of this segment on each transcript.

Table 1317 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R16276_PEA_1_T6	755	876

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster R16276_PEA_1_node_1 according to the present invention is supported by 37 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R16276_PEA_1_T6. Table 1318 below describes the starting and ending position of this segment on each transcript.

Table 1318 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R16276_PEA_1_T6	439	528

Segment cluster R16276_PEA_1_node_4 according to the present invention is supported by 38 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R16276_PEA_1_T6. Table 1319 below describes the starting and ending position of this segment on each transcript.

Table 1319 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R16276_PEA_1_T6	529	639

Segment cluster R16276_PEA_1_node_5 according to the present invention is supported by 37 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): R16276_PEA_1_T6. Table 1320 below describes the starting and ending position of this segment on each transcript.

Table 1320 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
R16276_PEA_1_T6	640	754

Variant protein alignment to the previously known protein:

Sequence name: NOV_HUMAN

Sequence documentation:

Alignment of: R16276_PEA_1_P7 x NOV_HUMAN ..

Alignment segment 1/1:

Quality: 1042.00

Escore: 0

Matching length: 103 Total length: 103

Matching Percent Similarity: 100.00 Matching Percent Identity: 99.03

Total Percent Similarity: 100.00 Total Percent Identity: 99.03

Gaps: 0

Alignment:

```
                .         .         .         .         .
      1 MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPGQCPATPPTC 50
        ||||||||||||||||||||||||||||||||||||||||:||||||||
      1 MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPGRCPATPPTC 50
                .         .         .         .         .
     51 APGVRAVLDGCSCCLVCARQRGESCSDLEPCDESSGLYCDRSADPSNQTG 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 APGVRAVLDGCSCCLVCARQRGESCSDLEPCDESSGLYCDRSADPSNQTG 100

    101 ICT                                               103
        |||
    101 ICT                                               103
```

Combined expression of 6 sequences H61775seg8, HUMGRP5E junc3-7, M85491Seg24, Z21368 junc17-21, HSSTROL3seg24 and Z25299seg20 in normal and cancerous lung tissues.

Expression of immunoglobulin superfamily, member 9, gastrin-releasing peptide, Ephrin type-B receptor 2 precursor, SUL1_HUMAN, Stromelysin-3 precursor (EC 3.4.24.-) (Matrix metalloproteinase-11) (MMP-11) (ST3) (SL-3) and Secretory leukocyte protease inhibitor Acid-stable proteinase inhibitor transcripts detectable by or according to H61775seg8 (SEQ ID NO: 1636), HUMGRP5E junc3-7 (SEQ ID NO: 1648), M85491Seg24 (SEQ ID NO: 1639), Z21368 junc17-21 (SEQ ID NO: 1642), HSSTROL3seg24 (SEQ ID NO: 1675) and Z25299seg20 amplicons (SEQ ID NO: 1669) and H61775seg8F (SEQ ID NO: 1634), H61775seg8R (SEQ ID NO: 1635), HUMGRP5E junc3-7F (SEQ ID NO: 1646), HUMGRP5E junc3-7R (SEQ ID NO: 1647), M85491Seg24F (SEQ ID NO: 1637), M85491Seg24R (SEQ ID NO: 1638), Z21368 junc17-21F (SEQ ID NO: 1640), Z21368 junc17-21R (SEQ ID NO: 1641), HSSTROL3seg24F (SEQ ID NO: 1673), HSSTROL3seg24R (SEQ ID NO: 1674), Z25299seg20F (SEQ ID NO: 1667), Z25299seg20R (SEQ ID NO: 1668) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), Ubiquitin (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicons was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample of each amplicon was then divided by the median of the quantities of the normal post-mortem (PM) samples detected for the same amplicon (Sample Nos. 47-50, 90-93, 96-99, Table 2,, "Tissue samplesin testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples. The reciprocal of this ratio was calculated for Z25299seg20 (SEQ ID NO: 1669), to obtain a value of fold down-regulation for each sample relative to median of the normal PM samples.

Figures 52-53 are histograms showing differential expression of the above-indicated transcripts in cancerous lung samples relative to the normal samples. The number and percentage

of samples that exhibit at least 5 fold differential of at least one of the sequences, out of the total number of samples tested is indicated in the bottom.

As is evident from Figures 52-53, differential expression of at least 5 fold in at least one of the sequences was found in 15 out of 15 adenocarcinoma samples, 14 out of 16 squamous cell carcinoma samples, 4 out of 4 large cell carcinoma samples and in 8 out of 8 small cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below. Threshold of 5 fold differential expression of at least one of the amplicons was found to differentiate between cancer and normal samples with P value of 7.82E-06 in adenocarcinoma, 2.63E-04 in squamous cell carcinoma, 8.24E-03 in large cell adenocarcinoma and 3.57E-04 in small cell carcinoma as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

DESCRIPTION FOR CLUSTER H53626

Cluster H53626 features 2 transcript(s) and 20 segment(s) of interest, the names for which are given in Tables 1321 and 1322, respectively, the sequences themselves are given at the end of the application.

Table 1321 - Transcripts of interest

Transcript Name	SEQ ID NO:
H53626_PEA_1_T15	16
H53626_PEA_1_T16	17

Table 1322 - Segments of interest

Segment Name	SEQ ID NO:

H53626_PEA_1_node_15	18
H53626_PEA_1_node_22	19
H53626_PEA_1_node_25	306
H53626_PEA_1_node_26	307
H53626_PEA_1_node_27	308
H53626_PEA_1_node_34	309
H53626_PEA_1_node_35	310
H53626_PEA_1_node_36	311
H53626_PEA_1_node_11	312
H53626_PEA_1_node_12	313
H53626_PEA_1_node_16	314
H53626_PEA_1_node_19	315
H53626_PEA_1_node_20	316
H53626_PEA_1_node_24	317
H53626_PEA_1_node_28	318
H53626_PEA_1_node_29	319
H53626_PEA_1_node_30	320
H53626_PEA_1_node_31	321
H53626_PEA_1_node_32	322
H53626_PEA_1_node_33	323

Table 1323 - Proteins of interest

Transcript Name	SEQ ID NO:
H53626_PEA_1_P4	324
H53626_PEA_1_P5	325

Cluster H53626 can be used as a diagnostic marker according to overexpression of transcripts of this cluster in cancer. Expression of such transcripts in normal tissues is also given according to the previously described methods. The term "number" in the right hand column of the table and the numbers on the y-axis of figure 76 below refer to weighted expression of ESTs

in each category, as "parts per million" (ratio of the expression of ESTs for a particular cluster to the expression of all ESTs in that category, according to parts per million).

Overall, the following results were obtained as shown with regard to the histograms in Figure 76 and Table *1324*. This cluster is overexpressed (at least at a minimum level) in the following pathological conditions: epithelial malignant tumors, a mixture of malignant tumors from different tissues and myosarcoma.

Table 1324 - Normal tissue distribution

Name of Tissue	Number
adrenal	4
bone	233
brain	33
colon	0
epithelial	12
general	17
head and neck	0
kidney	8
lung	25
breast	8
muscle	0
ovary	7
pancreas	10
prostate	8
skin	0
stomach	73
Thyroid	0
uterus	0

Table 1325 - P values and ratios for expression in cancerous tissue

1327

Name of Tissue	P1	P2	SP1	R3	SP2	R4
adrenal	6.4e-01	4.2e-01	2.1e-01	3.1	1.3e-02	4.1
bone	5.8e-01	8.1e-01	9.8e-01	0.3	1.0e+00	0.3
brain	2.2e-01	2.6e-01	8.1e-01	0.8	8.9e-01	0.6
colon	2.3e-01	1.4e-01	1.5e+00	1.2	4.6e-01	1.9
epithelial	8.3e-02	4.8e-03	6.4e-02	1.5	6.6e-08	4.1
general	2.4e-03	1.5e-05	1.1e-03	1.6	2.0e-12	3.1
head and neck	2.1e-01	3.3e-01	0.0e+00	0.0	0.0e+00	0.0
kidney	7.3e-01	5.8e-01	5.8e-01	1.3	5.7e-02	2.0
lung	8.3e-01	5.5e-01	7.9e-01	0.8	3.2e-02	2.1
breast	6.5e-01	2.7e-01	6.9e-01	1.2	7.8e-02	1.9
muscle	1.5e+00	2.9e-01	1.5e+00	1.0	3.5e-03	4.1
ovary	6.7e-01	5.6e-01	1.5e-01	1.7	7.0e-02	2.7
pancreas	2.3e-01	2.0e-01	3.9e-01	1.9	8.2e-02	2.3
prostate	9.0e-01	9.0e-01	6.7e-01	1.1	1.8e-01	1.9
skin	1.5e+00	4.4e-01	1.5e+00	1.0	6.4e-01	1.6
stomach	9.0e-01	3.4e-01	1.0e+00	0.3	6.1e-01	0.9
Thyroid	2.4e-01	2.4e-01	1.5e+00	1.1	1.5e+00	1.1
uterus	2.1e-01	2.4e-01	2.9e-01	2.5	2.6e-01	2.2

As noted above, contig H53626 features 2 transcript(s), which were listed in Table 1321 above. A description of each variant protein according to the present invention is now provided.

Variant protein H53626_PEA_1_P4 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H53626_PEA_1_T15. The alignment to the wild type protein is given at the end of the

application. A brief description of the relationship of the variant protein according to the present invention to the wild type protein is as follows:

Comparison report between H53626_PEA_1_P4 and wild type Q8N441 (SEQ ID NO:1699):

1.An isolated chimeric polypeptide encoding for H53626_PEA_1_P4, comprising a first amino acid sequence being at least 90 % homologous to MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPP LTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVV LDDISPGKESLGPDSSSGGQEDPASQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVAS GHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAINAT YKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGR HNSTIDVGGQKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFR SAFLTVLP corresponding to amino acids 1 - 357 of Q8N441, which also corresponds to amino acids 1 - 357 of H53626_PEA_1_P4, second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence GARLPRHATPCWCPDPPPGPGVPPTGWGPTLPSRAVLARSSAEGGQPRGTVSTAPGMG LGCSPGLCVGVPLPTSFPLALA corresponding to amino acids 358 - 437 of H53626_PEA_1_P4, and a third amino acid sequence being at least 90 % homologous to DPKPPGPPVASSSSATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGH RPPGTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKLYPKLY TDIHTHTHTHSHTHSHVEGKVHQHIHYQC corresponding to amino acids 358 - 504 of Q8N441, which also corresponds to amino acids 438 - 584 of H53626_PEA_1_P4, wherein said first, second and third amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for an edge portion of H53626_PEA_1_P4, comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence encoding for GARLPRHATPCWCPDPPPGPGVPPTGWGPTLPSRAVLARSSAEGGQPRGTVSTAPGMG LGCSPGLCVGVPLPTSFPLALA, corresponding to H53626_PEA_1_P4.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: membrane. The protein localization is believed to be membrane because although both signal-peptide prediction programs agree that this protein has a signal peptide, both trans-membrane region prediction programs predict that this protein has a trans-membrane region downstream of this signal peptide..

Variant protein H53626_PEA_1_P4 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table *1326*, (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H53626_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1326 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
193	R -> L	Yes
300	G ->	No
319	Y -> H	No
442	P -> Q	Yes
504	R -> L	Yes
521	G ->	No
544	P -> L	Yes
573	E -> G	No

Variant protein H53626_PEA_1_P4 is encoded by the following transcript(s): H53626_PEA_1_T15, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H53626_PEA_1_T15 is shown in bold; this coding portion starts at position 17 and ends at position 1771. The transcript also has the following SNPs as listed in Table *1327* (given according to their position on the nucleotide sequence, with the alternative

nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H53626_PEA_1_P4 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1327 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
76	G -> A	Yes
340	G -> T	No
1647	C -> T	Yes
1734	A -> G	No
1797	G ->	No
1948	A -> G	Yes
2193	C -> T	Yes
2308	C -> T	Yes
2333	C -> G	Yes
2648	C -> T	Yes
2649	G -> A	Yes
2765	C -> T	Yes
594	G -> T	Yes
2972	G -> A	Yes
3027	C -> G	Yes
907	T -> C	Yes
916	C ->	No
971	T -> C	No
1135	G -> A	Yes
1341	C -> A	Yes
1527	G -> T	Yes
1579	C ->	No

Variant protein H53626_PEA_1_P5 according to the present invention has an amino acid sequence as given at the end of the application; it is encoded by transcript(s) H53626_PEA_1_T16. The alignment to the wild type protein is given at the end of the application. A brief description of the relationship of the variant protein according to the present invention to the wild type protein is as follows:

Comparison report between H53626_PEA_1_P5 and wild type Q9H4D7 (SEQ ID NO:1700):

1.An isolated chimeric polypeptide encoding for H53626_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPP LTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVV LDDISPGKESLGPDSSSGGQEDPASQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVAS GHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAINAT YKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCK corresponding to amino acids 1 - 269 of Q9H4D7, which also corresponds to amino acids 1 - 269 of H53626_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TQNRQGHLWPPRPRPLACRGPWSSASQPALSSSWAPCSCGFARPRRSRAPPRLPLPCLG TARRGRPATAAETRTFPRWPPSALALVWGCVRSMGLRQPPSTYWAQAQLLALSCTPNS TQTSTHTHTHTLTHTHTWRARSTSTSTISARRHRICSGHGGAGQTGRLGGWRTELQTKA GDPWRGGMASTPGSLCVRHSPWTHTHRHTHYLDACMHTHARTRAP corresponding to amino acids 270 - 490 of H53626_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of H53626_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TQNRQGHLWPPRPRPLACRGPWSSASQPALSSSWAPCSCGFARPRRSRAPPRLPLPCLG TARRGRPATAAETRTFPRWPPSALALVWGCVRSMGLRQPPSTYWAQAQLLALSCTPNS TQTSTHTHTHTLTHTHTWRARSTSTSTISARRHRICSGHGGAGQTGRLGGWRTELQTKA

GDPWRGGMASTPGSLCVRHSPWTHTHRHTHYLDACMHTHARTRAP in H53626_PEA_1_P5.

Comparison report between H53626_PEA_1_P5 and wild type Q8N441:

1.An isolated chimeric polypeptide encoding for H53626_PEA_1_P5, comprising a first amino acid sequence being at least 90 % homologous to MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPP LTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVV LDDISPGKESLGPDSSSGGQEDPASQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVAS GHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAINAT YKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCK corresponding to amino acids 1 - 269 of Q8N441, which also corresponds to amino acids 1 - 269 of H53626_PEA_1_P5, and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence TQNRQGHLWPPRPRPLACRGPWSSASQPALSSSWAPCSCGFARPRRSRAPPRLPLPCLG TARRGRPATAAETRTFPRWPPSALALVWGCVRSMGLRQPPSTYWAQAQLLALSCTPNS TQTSTHTHTHTLTHTHTWRARSTSTSTISARRHRICSGHGGAGQTGRLGGWRTELQTKA GDPWRGGMASTP GSLCVRHSPWTHTHRHTHYLDACMHTHARTRAP corresponding to amino acids 270 - 490 of H53626_PEA_1_P5, wherein said first and second amino acid sequences are contiguous and in a sequential order.

2.An isolated polypeptide encoding for a tail of H53626_PEA_1_P5, comprising a polypeptide being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence TQNRQGHLWPPRPRPLACRGPWSSASQPALSSSWAPCSCGFARPRRSRAPPRLPLPCLG TARRGRPATAAETRTFPRWPPSALALVWGCVRSMGLRQPPSTYWAQAQLLALSCTPNS TQTSTHTHTHTLTHTHTWRARSTSTSTISARRHRICSGHGGAGQTGRLGGWRTELQTKA GDPWRGGMASTPGSLCVRHSPWTHTHRHTHYLDACMHTHARTRAP in H53626_PEA_1_P5.

The location of the variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be located as follows with regard to the cell: secreted. The protein localization is believed to be secreted because both signal-peptide prediction programs predict that this protein has a signal peptide, and neither trans-membrane region prediction program predicts that this protein has a trans-membrane region..

Variant protein H53626_PEA_1_P5 also has the following non-silent SNPs (Single Nucleotide Polymorphisms) as listed in Table *1328* (given according to their position(s) on the amino acid sequence, with the alternative amino acid(s) listed; the last column indicates whether the SNP is known or not; the presence of known SNPs in variant protein H53626_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1328 - Amino acid mutations

SNP position(s) on amino acid sequence	Alternative amino acid(s)	Previously known SNP?
193	R -> L	Yes
274	Q -> K	Yes
336	A -> S	Yes
353	A ->	No
376	Q -> *	Yes
405	R -> G	No
426	G ->	No
476	Y -> C	Yes

Variant protein H53626_PEA_1_P5 is encoded by the following transcript(s): H53626_PEA_1_T16, for which the sequence(s) is/are given at the end of the application. The coding portion of transcript H53626_PEA_1_T16 is shown in bold; this coding portion starts at position 17 and ends at position 1489. The transcript also has the following SNPs as listed in Table *1329* (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed; the last column indicates whether the SNP is known or not; the presence of

known SNPs in variant protein H53626_PEA_1_P5 sequence provides support for the deduced sequence of this variant protein according to the present invention).

Table 1329 - Nucleic acid SNPs

SNP position on nucleotide sequence	Alternative nucleic acid	Previously known SNP?
76	G -> A	Yes
340	G -> T	No
1688	C -> T	Yes
1803	C -> T	Yes
1828	C -> G	Yes
2143	C -> T	Yes
2144	G -> A	Yes
2260	C -> T	Yes
2467	G -> A	Yes
2522	C -> G	Yes
594	G -> T	Yes
836	C -> A	Yes
1022	G -> T	Yes
1074	C ->	No
1142	C -> T	Yes
1229	A -> G	No
1292	G ->	No
1443	A -> G	Yes

As noted above, cluster H53626 features 20 segment(s), which were listed in Table 2 above and for which the sequence(s) are given at the end of the application. These segment(s) are portions of nucleic acid sequence(s) which are described herein separately because they are

of particular interest. A description of each segment according to the present invention is now provided.

Segment cluster H53626_PEA_1_node_15 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1330* below describes the starting and ending position of this segment on each transcript.

Table 1330 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	96	343
H53626_PEA_1_T16	96	343

Segment cluster H53626_PEA_1_node_22 according to the present invention is supported by 42 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1332* below describes the starting and ending position of this segment on each transcript.

Table 1332 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	450	734
H53626_PEA_1_T16	450	734

Segment cluster H53626_PEA_1_node_25 according to the present invention is supported by 41 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15. Table *1334* below describes the starting and ending position of this segment on each transcript.

Table 1334 - Segment location on transcripts

1336

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	824	1088

Segment cluster H53626_PEA_1_node_26 according to the present invention is supported by 5 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15. Table *1336* below describes the starting and ending position of this segment on each transcript.

Table 1336 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	1089	1328

Segment cluster H53626_PEA_1_node_27 according to the present invention is supported by 106 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1338* below describes the starting and ending position of this segment on each transcript.

Table 1338 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	1329	2228
H53626_PEA_1_T16	824	1723

Segment cluster H53626_PEA_1_node_34 according to the present invention is supported by 121 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1340* below describes the starting and ending position of this segment on each transcript.

Table 1340- Segment location on transcripts

Transcript name	Segment starting position	Segment ending position

H53626_PEA_1_T15	2507	2977
H53626_PEA_1_T16	2002	2472

Segment cluster H53626_PEA_1_node_35 according to the present invention is supported by 85 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1342* below describes the starting and ending position of this segment on each transcript.

Table 1342 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2978	3148
H53626_PEA_1_T16	2473	2643

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment, shown in Table *1343*.

Table 1343 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
NA		

Segment cluster H53626_PEA_1_node_36 according to the present invention is supported by 69 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1344* below describes the starting and ending position of this segment on each transcript.

Table 1344 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	3149	3322
H53626_PEA_1_T16	2644	2817

Microarray (chip) data is also available for this segment as follows. As described above with regard to the cluster itself, various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer. The following oligonucleotides were found to hit this segment, shown in Table *13455*.

Table 1345 - Oligonucleotides related to this segment

Oligonucleotide name	Overexpressed in cancers	Chip reference
NA		

According to an optional embodiment of the present invention, short segments related to the above cluster are also provided. These segments are up to about 120 bp in length, and so are included in a separate description.

Segment cluster H53626_PEA_1_node_11 according to the present invention is supported by 12 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1346* below describes the starting and ending position of this segment on each transcript.

Table 1346 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	1	55
H53626_PEA_1_T16	1	55

Segment cluster H53626_PEA_1_node_12 according to the present invention is supported by 11 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1347* below describes the starting and ending position of this segment on each transcript.

Table 1347 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	56	95
H53626_PEA_1_T16	56	95

Segment cluster H53626_PEA_1_node_16 according to the present invention can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1348* below describes the starting and ending position of this segment on each transcript.

Table 1348 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	344	368
H53626_PEA_1_T16	344	368

Segment cluster H53626_PEA_1_node_19 according to the present invention is supported by 25 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1349* below describes the starting and ending position of this segment on each transcript.

Table 1349 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	369	419
H53626_PEA_1_T16	369	419

Segment cluster H53626_PEA_1_node_20 according to the present invention is supported by 27 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1350* below describes the starting and ending position of this segment on each transcript.

Table 1350 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	420	449
H53626_PEA_1_T16	420	449

Segment cluster H53626_PEA_1_node_24 according to the present invention is supported by 34 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1351* below describes the starting and ending position of this segment on each transcript.

Table 1351 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	735	823
H53626_PEA_1_T16	735	823

Segment cluster H53626_PEA_1_node_28 according to the present invention is supported by 66 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1352* below describes the starting and ending position of this segment on each transcript.

Table 1352 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2229	2306
H53626_PEA_1_T16	1724	1801

Segment cluster H53626_PEA_1_node_29 according to the present invention is supported by 73 libraries. The number of libraries was determined as previously described. This segment

can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1353* below describes the starting and ending position of this segment on each transcript.

Table 1353 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2307	2396
H53626_PEA_1_T16	1802	1891

Segment cluster H53626_PEA_1_node_30 according to the present invention is supported by 71 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1354* below describes the starting and ending position of this segment on each transcript.

Table 1354 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2397	2442
H53626_PEA_1_T16	1892	1937

Segment cluster H53626_PEA_1_node_31 according to the present invention is supported by 67 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1355* below describes the starting and ending position of this segment on each transcript.

Table 1355 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2443	2469
H53626_PEA_1_T16	1938	1964

Segment cluster H53626_PEA_1_node_32 according to the present invention is supported by 65 libraries. The number of libraries was determined as previously described. This segment can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1356* below describes the starting and ending position of this segment on each transcript.

Table 1356 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2470	2498
H53626_PEA_1_T16	1965	1993

Segment cluster H53626_PEA_1_node_33 according to the present invention can be found in the following transcript(s): H53626_PEA_1_T15 and H53626_PEA_1_T16. Table *1357* below describes the starting and ending position of this segment on each transcript.

Table 1357 - Segment location on transcripts

Transcript name	Segment starting position	Segment ending position
H53626_PEA_1_T15	2499	2506
H53626_PEA_1_T16	1994	2001

Variant protein alignment to the previously known protein:

```
Sequence name: /tmp/K1Mec2ReKO/eg1EUS2AXY:Q8N441

Sequence documentation:

Alignment of: H53626_PEA_1_P4 x Q8N441    ..

Alignment segment 1/1:
```

1343

Quality: 4882.00

Escore: 0

Matching length: 504 Total length: 584

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 86.30 Total Percent Identity: 86.30

Gaps: 1

Alignment:

```
                .         .         .         .         .
      1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
                .         .         .         .         .
     51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
                .         .         .         .         .
    101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
                .         .         .         .         .
    151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
                .         .         .         .         .
    201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
```

```
                .         .         .         .         .
251 THPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGRHNSTIDVGG 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
251 THPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGRHNSTIDVGG 300
                .         .         .         .         .
301 QKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRS 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
301 QKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRS 350
                .         .         .         .         .
351 AFLTVLPGARLPRHATPCWCPDPPPGPGVPPTGWGPTLPSRAVLARSSAE 400
    |||||||
351 AFLTVLP..................................... 357
                .         .         .         .         .
401 GGQPRGTVSTAPGMGLGCSPGLCVGVPLPTSFPLALADPKPPGPPVASSS 450
                                                ||||||||||||
358 ..................................DPKPPGPPVASSS 370
                .         .         .         .         .
451 SATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPP 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
371 SATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPP 420
                .         .         .         .         .
501 GTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKL 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
421 GTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKL 470
                .         .         .
551 YPKLYTDIHTHTHTHSHTHSHVEGKVHQHIHYQC                  584
    ||||||||||||||||||||||||||||||||
471 YPKLYTDIHTHTHTHSHTHSHVEGKVHQHIHYQC                  504
```

Sequence name: /tmp/oSUZaRW3WK/oSh3fN5Zt0:Q9H4D7

Sequence documentation:

Alignment of: H53626_PEA_1_P5 x Q9H4D7 ..

Alignment segment 1/1:

Quality: 2644.00
Escore: 0
Matching length: 269 Total length: 269
Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00
Total Percent Similarity: 100.00 Total Percent Identity: 100.00
Gaps: 0

Alignment:

```
                .         .         .         .         .
        1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
        1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
                .         .         .         .         .
       51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
                .         .         .         .         .
```

```
101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
              .         .         .         .         .
151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
              .         .         .         .         .
201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||
201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
              .
251 THPVNTTVDFGGTTSFQCK                                 269
    |||||||||||||||||||
251 THPVNTTVDFGGTTSFQCK                                 269
```

Sequence name: /tmp/oSUZaRW3WK/oSh3fN5Zt0:Q8N441

Sequence documentation:

Alignment of: H53626_PEA_1_P5 x Q8N441 ..

Alignment segment 1/1:

Quality: 2644.00

Escore: 0

Matching length: 269 Total length: 269

Matching Percent Similarity: 100.00 Matching Percent Identity: 100.00

Total Percent Similarity: 100.00 Total Percent Identity: 100.00

Gaps: 0

Alignment:

```
                .         .         .         .         .
  1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQ 50
                .         .         .         .         .
 51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 CPVEGDPPPLTMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCK 100
                .         .         .         .         .
101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 ATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFT 150
                .         .         .         .         .
151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 QPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPR 200
                .         .         .         .         .
201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 KKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTG 250
                .
251 THPVNTTVDFGGTTSFQCK                                 269
```

1348

||||||||||||||||||||

251 THPVNTTVDFGGTTSFQCK 269

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) *H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 junc24-27F1R3 in normal and cancerous lung tissues.*

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1)transcripts detectable by or according to junc24-27, H53626 junc24-27F1R3 amplicon (SEQ ID NO: 1690) and H53626 junc24-27F1 (SEQ ID NO: 1688) and H53626 junc24-27R3 (SEQ ID NO: 1689) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 74 is a histogram showing over expression of the above-indicated Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts in cancerous lung samples relative to the normal samples.

As is evident from Figure 74, the expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts detectable by the above amplicon(s) was higher in several cancer samples than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 5 fold was found in 7 out of 15 adenocarcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: H53626 junc24-27F1 forward primer; and H53626 junc24-27R3 reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: H53626 junc24-27F1R3.

Forward primer (SEQ ID NO: 1688): GTCCTTCCAGTGCAAGACCCA

Reverse primer(SEQ ID NO: 1689): TGGGCCTGGCAAAGCC

Amplicon (SEQ ID NO: 1690):

GTCCTTCCAGTGCAAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTC
GGCCACTAGCCTGCCGTGGCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCAT
CCTGGGCACCCTGCTCCTGTGGCTTTGCCAGGCCCA

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) *H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 seg25 in normal and cancerous lung tissues.*

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts detectable by or according to seg25, H53626 seg25 amplicon (SEQ ID NO: 1693) and H53626 seg25F (SEQ ID NO: 1691) and H53626 seg25R (SEQ ID NO: 1692) primers was measured by real time PCR. In parallel the expression of four housekeeping genes –PBGD (GenBank Accession No. BC019323; amplicon – PBGD-amplicon, SEQ ID NO:334), HPRT1 (GenBank Accession No. NM_000194; amplicon – HPRT1-amplicon, SEQ ID NO:1297), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331), was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-

mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

As is evident from Figure 75, the expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts detectable by the above amplicon(s) was higher in a few cancer samples than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2). Notably an over-expression of at least 5 fold was found in 3 out of 15 adenocarcinoma samples.

Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: H53626 seg25F forward primer; and H53626 seg25R reverse primer.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: H53626 seg25.
Forward primer (SEQ ID NO: 1691);CCGACGGCTCCTACCTCAA
Reverse primer (SEQ ID NO: 1692): GGAAGCTGTAGCCCATGGTGT
Amplicon (SEQ ID NO: 1693):
CCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCGTGCCCGCCAGGACGATGCG
GGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTACAGCTTCC

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 seg25 in different normal tissues.

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts detectable by or according to H53626 seg25 amplicon (SEQ ID NO: 1693) and H53626 seg25F (SEQ ID NO: 1691) and H53626 seg25R (SEQ ID NO: 1692) was measured by real time PCR. In parallel the expression of four housekeeping genes: RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No.

NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the lung samples (Sample Nos. 15-17 Table 3 above), to obtain a value of relative expression of each sample relative to median of the lung samples.

Forward primer (SEQ ID NO: 1691);CCGACGGCTCCTACCTCAA
Reverse primer (SEQ ID NO: 1692): GGAAGCTGTAGCCCATGGTGT
Amplicon (SEQ ID NO: 1693):
CCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCGTGCCCGCCAGGACGATGCG
GGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTACAGCTTCC

The results are demonstrated in Figure 77, showing the expression of *of* Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 seg25 in different normal tissues.

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts which are detectable by amplicon as depicted in sequence name H53626 junc24-27F1R3 in different normal tissues

Expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) transcripts detectable by or according to H53626 junc24-27F1R3 amplicon (SEQ ID NO: 1690) and H53626 junc24-27F1 (SEQ ID NO: 1688) and H53626 junc24-27R3 (SEQ ID NO: 1689) was measured by real time PCR. In parallel the expression of four housekeeping genes – RPL19 (GenBank Accession No. NM_000981; RPL19 amplicon, SEQ ID NO:1630), TATA box (GenBank Accession No. NM_003194; TATA amplicon, SEQ ID NO:1633; primers SEQ ID NOs 1631 and 1632), UBC (GenBank Accession No. BC000449; amplicon – Ubiquitin-amplicon, SEQ ID NO:328) and SDHA (GenBank Accession No. NM_004168; amplicon – SDHA-amplicon, SEQ ID NO:331) was measured similarly. For each RT sample, the

expression of the above amplicon was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the lung samples (Sample Nos. 15-17 Table 3 above), to obtain a value of relative expression of each sample relative to median of the lung samples.

Forward primer (SEQ ID NO: 1688): GTCCTTCCAGTGCAAGACCCA
Reverse primer(SEQ ID NO: 1689): TGGGCCTGGCAAAGCC
Amplicon (SEQ ID NO: 1690):
GTCCTTCCAGTGCAAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTC
GGCCACTAGCCTGCCGTGGCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCAT
CCTGGGCACCCTGCTCCTGTGGCTTTGCCAGGCCCA

The results are demonstrated in Figure 78, showing the expression of Homo sapiens fibroblast growth factor receptor-like 1 (FGFRL1) H53626 transcripts, which are detectable by amplicon as depicted in sequence name H53626 *junc24-27F1R3* in different normal tissues.

Expression of trophinin associated protein (tastin) [T86235] transcripts which are detectable by amplicon as depicted in SEQ ID NO:1480 in normal and cancerous lung tissues

Expression of trophinin associated protein (tastin) transcripts detectable by SEQ ID NO:1480 (e.g., variant no. 23-26 31, 32- represented by SEQ IDs 1485-1488, 1609, 1610) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO:1477), was measured similarly. For each RT sample, the expression of SEQ ID NO:1480 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 54a is a histogram showing over expression of the above-indicated trophinin associated protein (tastin) transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 5 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 54a, the expression of trophinin associated protein (tastin) transcripts detectable by SEQ ID NO:1480 in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99 Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 6 out of 15 adenocarcinoma samples, 8 out of 16 squamous cell carcinoma samples, 2 out of 4 large cell carcinoma samples and in 8 out of 8 small cells carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of trophinin associated protein (tastin) transcripts detectable by SEQ ID NO:1480 in lung cancer samples versus the normal lung samples was determined by T test as 1.61E-04.

Threshold of 5 fold overexpression was found to differentiate between cancer and normal samples with P value of 1.49E-02 as checked by exact fisher test. The above values demonstrate statistical significance of the results.

According to the present invention, trophinin associated protein (tastin) is a non-limiting example of a marker for diagnosing lung cancer. The trophinin associated protein (tastin) marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to trophinin associated protein (tastin) as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: trophinin associated protein (tastin)-TAA-seg 44-forward primer (SEQ ID NO: 1478):

AGACTCCAACCCACAGCCC; and trophinin associated protein (tastin) - TAA-seg 44-Reverse primer (SEQ ID NO: 1479): CAGCTCAGCCAACCTTGCA.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: trophinin associated protein (tastin) amplicon, SEQ ID NO: 1480:
AGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGG
AACAGGGACAGGGGGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTT
GGCTACAGTGCAAGGTTGGCTGAGCTG

According to other preferred embodiments of the present invention, trophinin associated protein (tastin) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, trophinin associated protein (tastin) splice variants, as depicted in SEQ ID NO: 1485-1488, 1609, 1610 (e.g., variant no. 23-26, 31, 32), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of trophinin associated protein (tastin) comprises segment_TAA-44 – SEQ ID NO: 1507. Also optionally and more preferably, any suitable method may be used for detecting a fragment such as trophinin associated protein (tastin) _segment_ TAA-44 – SEQ ID no 1507 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to *trophinin associated protein (tastin)* as described above, including but not limited to SEQ ID NOs: 1492-1501, 1612. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequences of these proteins that are depicted in SEQ ID Nos: 1508-1511, 1613. The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to *trophinin associated protein (tastin)* as described above, optionally for any application.

Expression of trophinin associated protein (tastin) [T86235] transcripts which are detectable by oligonucleotides as depicted in SEQ ID NOs:1512-1514 in normal and cancerous lung tissues

Expression of *trophinin associated protein (tastin) [T86235]* transcripts detectable by oligonucleotides SEQ ID NOs: *1512-1514* (e.g., variants no. 8-10, 22, 23, 26, 27, 29-31, 33 – represented by SEQ IDs 1481-1485, 1488-1491, 1609, 1611) was measured with oligonucleotide-based micro-arrays. The segments detected by the above oligonucleotides as depicted in SEQ ID NOs: *1512-1514* are for example nucleotide sequences as depicted in SEQ IDs 1503, 1504, 1506.

The results of image intensities for each feature were normalized according to the ninetieth percentile of the image intensities of all the features on the chip. Then, feature image intensities for replicates of the same oligonucleotide on the chip and replicates of the same sample were averaged. Outlying results were discarded.

For every oligonucleotide (SEQ ID NOs: *1512-1514*) the averaged intensity determined for every sample was divided by the averaged intensity of all the normal samples (Sample Nos. 48,50, 90-92, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold

up-regulation for each sample relative to the averaged normal samples. These data are presented in a histogram in Figure 54b. As is evident from Figure 54b, the expression of *trophinin associated protein (tastin) [T86235]* transcripts detectable with oligonucleotides according to SEQ ID NOs: *1512-1514* in cancer samples was significantly higher than in the normal samples.

According to the present invention, trophinin associated protein (tastin) is a non-limiting example of a marker for diagnosing lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to trophinin associated protein (tastin) as previously defined is also encompassed within the present invention. Oligonucleotides are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following oligonucleotides were used as a non-limiting illustrative example only of a suitable oligonucleotides: SEQ ID NOs: *1512-1514*

SEQ ID *1512*:

CATGGTAACACGGCCTCCATGGCTGAGTAGGGGACTAGGAAGGGTAAAAG

SEQ ID 1513:

TGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGG

SEQ ID 1514:

TGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGC

According to other preferred embodiments of the present invention, trophinin associated protein (tastin) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, trophinin associated protein (tastin) splice variants, as depicted in SEQ ID NO: 1481-1485, 1488-1491, 1609, 1611 (e.g., variant no. 8-10, 22, 23, 26, 27, 29-31, 33), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of trophinin associated protein (tastin) comprises segment_TAA-14, 35 and 42 – SEQ ID no. 1503, 1504, 1506 . Also optionally and more preferably, any suitable method may be used for detecting a fragment such as trophinin associated protein (tastin) _segment_TAA-14, 35 and 42 – SEQ ID NOs 1503, 1504 and 1506 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of

specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments of the present invention, trophinin associated protein (tastin) splice variants containing the unique segments as depicted in SEQ ID Nos 1502 and 1505, for example as these included in variants 9 and 29 (SEQ ID NOs: 1482 and 1490, respectively), are useful as biomarkers for detecting lung cancer.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to *trophinin associated protein (tastin)* as described above, optionally for any application.

Expression of Homeo box C10 (HOXC10) [N31842] transcripts which are detectable by amplicon as depicted in SEQ ID NO:1517 in normal and cancerous lung tissues

Expression of Homeo box C10 (HOXC10) transcripts detectable by SEQ ID NO: *1517* (e.g., variant no. 3, represented by SEQ ID 1519) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:3), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:9) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO:1477), was measured similarly. For each RT sample, the expression of SEQ ID NO:1517 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 55 is a histogram showing over expression of the above-indicated Homeo box C10 (HOXC10) transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 20 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 55, the expression of Homeo box C10 (HOXC10) transcripts detectable by SEQ ID NO: 1517 in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing

panel"). Notably an over-expression of at least 20 fold was found in 6 out of 15 adenocarcinoma samples, 9 out of 16 squamous cell carcinoma samples, and in 3 out of 4 large cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Homeo box C10 (HOXC10) transcripts detectable by SEQ ID NO: 1517 in lung cancer samples versus the normal lung samples was determined by T test as 4.43E-03.

Threshold of 20 fold overexpression was found to differentiate between cancer and normal samples with P value of 2.88E-02 as checked by exact fisher test. The above values demonstrate statistical significance of the results.

According to the present invention, Homeo box C10 (HOXC10) is a non-limiting example of a marker for diagnosing lung cancer. The Homeo box C10 (HOXC10) marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to Homeo box C10 (HOXC10) as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Homeo box C10 (HOXC10) - forward primer (SEQ ID NO: 1515): GCGAAACGCGATTTGTTGTT; and Homeo box C10 (HOXC10) -Reverse primer (SEQ ID NO:1516): CATCTGGAGGAGGGAGGGA.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon

was obtained as a non-limiting illustrative example only of a suitable amplicon: Homeo box C10 (HOXC10) amplicon (SEQ ID NO:1517):

GCGAAACGCGATTTGTTGTTTGTGGGTCTGATTTGTGCGTGCGGCTTGGGCTCCTGC
GGCTTTTGGCTCGGCCGGGGGCCTTGGGCAGCGAGGCTGGAGCCGGAAGAGGTGG
AGGTGAAGGGCTGCCCGCCACGTCCCTCCCTCCTCCAGATG .

According to other preferred embodiments of the present invention, Homeo box C10 (HOXC10) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, Homeo box C10 (HOXC10) splice variants, as depicted in SEQ ID NO:54 (e.g., variant no. 3), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of Homeo box C10 (HOXC10) comprises segment_TAA-seg 6 (SEQ ID NO: 1526). Also optionally and more preferably, any suitable method may be used for detecting a fragment such as Homeo box C10 (HOXC10) _segment_ TAA-seg 6 (SEQ ID NO:1526) for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments of the present invention, *Homeo box C10 (HOXC10)* splice variants containing the unique segments as depicted in SEQ ID NOs: 1524 and 1525, for example transcripts as depicted in SEQ ID NO: 1515, 1519 and 1520, comprise a biomarker for detecting lung cancer.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to *trophinin associated protein (tastin)* as described above, including but not limited to SEQ ID NOs: 1521 and 1522. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequence of the protein SEQ ID NO: 1522, as depicted in SEQ ID NO:1523. The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to *trophinin associated protein (tastin)* as described above, optionally for any application.

Expression of Nucleolar protein 4 (NOL4)- [T06014] transcripts which are detectable by amplicon as depicted in SEQ IDs NO:1529 in normal and cancerous lung tissues

Expression of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NOs:1529 (e.g., variant no. 3, 11 and 12, represented by SEQ IDs 1533, 1537, 1538) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO:1477), was measured similarly. For each RT sample, the expression of SEQ ID NO:1529 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, above, "Tissue samples in testing panel"), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figures 56a and b are histograms showing over expression of the above-indicated Nucleolar protein 4 (NOL4) transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 200 fold or 6 fold over-

expression, out of the total number of samples tested is indicated in the bottom of figures 56a and 56b respectively.

As is evident from Figure 56a, the expression of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO: 1529 in the samples originate from small cell carcinoma of the lung was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 200 fold was found in 8 out of 8 small cell carcinoma samples. As is evident from Figures 56b, over expression of at least 6 fold was observed also in 2 out of 15 adenocarcinoma samples, 3 out of 16 squamous cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO:1529 in lung cancer samples versus the normal lung samples was determined by T test as 1.36E-02.

Threshold of 6 fold overexpression was found to differentiate between cancer and normal samples with P value of 2.52E-02 as checked by exact fisher test.

The P value for the difference in the expression levels of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO:1529 in lung small cell carcinoma samples versus the normal lung samples was determined by T test as 3.86E-03.

Threshold of 200 fold overexpression was found to differentiate between small cell carcinoma and normal lung samples with P value of 7.94E-06 as checked by exact fisher test.
The above values demonstrate statistical significance of the results.

According to the present invention, Nucleolar protein 4 (NOL4) is a non-limiting example of a marker for diagnosing lung cancer. The Nucleolar protein 4 (NOL4) marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to Nucleolar protein 4 (NOL4) as previously defined is also

encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Nucleolar protein 4 (NOL4)-TAA-seg1-forward primer (SEQ ID NO:1527): CTCGCTCCCTTGCTCACAC; and Nucleolar protein 4 (NOL4)-TAA-seg1 -Reverse primer (SEQ ID NO:1528): AAAGGGAAAGCGGGATGTTT.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Nucleolar protein 4 (NOL4) amplicon (SEQ ID NO:1529):
CTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACT
GACCATTTTGCAAGTGTCAGGACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCT
TTCCCTTT.

According to other preferred embodiments of the present invention, Nucleolar protein 4 (NOL4) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, Nucleolar protein 4 (NOL4) splice variants, as depicted in SEQ ID NO:1529 (e.g., variants nos. 3, 11 and 12), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of Nucleolar protein 4 (NOL4) comprises segment_TAA-seg-1 (SEQ ID NO: 1552). Also optionally and more preferably, any suitable method may be used for detecting a fragment such as Nucleolar protein 4 (NOL4)_segment_ TAA-seg-1 (SEQ ID NO: 1552) for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments of the present invention, Nucleolar protein 4 (NOL4) splice variants containing the unique segments as depicted in SEQ ID NOs: 1554 and 1555, for example transcripts as depicted in SEQ ID NOs: 1534-1536 and 1539-1541, comprises a biomarker for detecting lung cancer.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid

sequence corresponding to Nucleolar protein 4 (NOL4) as described above, including but not limited to SEQ ID Nos: 1542, 1547 and 1543; 1548, 1545, 1546, and 1549-1551. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequence of the protein SEQ ID NO: 1543, 1546, 1549 as depicted in SEQ ID NO:1544.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to Nucleolar protein 4 (NOL4) as described above, optionally for any application.

Expression of Nucleolar protein 4 (NOL4)- [T06014] transcripts which are detectable by amplicon as depicted in SEQ IDs NO:1532 in normal and cancerous lung tissues

Expression of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NOs:1532 (e.g., variant no. 3, 11 and 12, represented by SEQ IDs 1533, 1537, 1538) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO: 1481), was measured similarly. For each RT sample, the expression of SEQ ID NO:1532 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figures 57a and b are histograms showing over expression of the above-indicated Nucleolar protein 4 (NOL4) transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 400 fold or 6 fold over-

expression, out of the total number of samples tested is indicated in the bottom of figures 57a and b respectively.

As is evident from Figure 57a, the expression of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO:1532 in the samples originate from small cell carcinoma of the lung was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 400 fold was found in 8 out of 8 small cell carcinoma samples. As is evident from Figure 4b, over expression of at least 6 fold was observed also in 4 out of 15 adenocarcinoma samples, 3 out of 16 squamous cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO:1532 in lung cancer samples versus the normal lung samples was determined by T test as 1.70E-02.

Threshold of 6 fold overexpression was found to differentiate between cancer and normal samples with P value of 1.80E-02 as checked by exact fisher test.

The P value for the difference in the expression levels of Nucleolar protein 4 (NOL4) transcripts detectable by SEQ ID NO:1532 in lung small cell carcinoma samples versus the normal lung samples was determined by T test as 7.08E-03.

Threshold of 400 fold overexpression was found to differentiate between small cell carcinoma and normal lung samples with P value of 1.03E-04 as checked by exact fisher test.
The above values demonstrate statistical significance of the results.

According to the present invention, Nucleolar protein 4 (NOL4) is a non-limiting example of a marker for diagnosing lung cancer. The Nucleolar protein 4 (NOL4) marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to Nucleolar protein 4 (NOL4) as previously defined is also

encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Nucleolar protein 4 (NOL4) –TAA-seg 3-forward primer (SEQ ID NO: 1530): ACATCCCCCTGGAACGGAT; and Nucleolar protein 4 (NOL4)-TAA-seg 3-Reverse primer (SEQ ID NO:1531): CAGAAATTAGCAAAGCATTGATGG.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Nucleolar protein 4 (NOL4) amplicon (SEQ ID NO: 1532): ACATCCCCCTGGAACGGATATCTGTTTGGGGCACTAC AATCTATCCTGTAGAACTAT GGCCAAATCTCCATCAATGCTTTGCTAATTTCTG .

According to other preferred embodiments of the present invention, Nucleolar protein 4 (NOL4) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, Nucleolar protein 4 (NOL4) splice variants, as depicted in SEQ ID NO:1533, 1537, 1538 (e.g., variants nos. 3, 11, 12), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of Nucleolar protein 4 (NOL4) comprises segment_TAA-seg-3 (SEQ ID NO: 1553). Also optionally and more preferably, any suitable method may be used for detecting a fragment such as Nucleolar protein 4 (NOL4)_segment_ TAA-seg-3 (SEQ ID NO: 1553) for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to Nucleolar protein 4 (NOL4) as described above, including but not limited to SEQ ID NOs: SEQ ID Nos: 1542, 1547 and 1548. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to Nucleolar protein 4 (NOL4) as described above, optionally for any application.

Expression of AA281370 transcripts which are detectable by amplicon as depicted in SEQ ID NO:1558 in normal and cancerous lung tissues

AA281370 gene was identified by a computational process described above as over expressed in lung cancer. The AA281370 encoded proteins (SEQ ID NO: 1563, 1564) contain several WD40 domains, which are found in a number of eukaryotic proteins that cover a wide variety of functions, including adaptor/regulatory modules in signal transduction, pre-mRNA processing and cytoskeleton assembly. As is demonstrated in Figure 63, the WD40 domain region of AA281370 encoded protein, depicted in SEQ ID NO: 1564, has several similarities that might suggest involvement in signal transduction MAPK pathway. For example, the region of the AA281370 polypeptide SEQ ID NO: 1564 located between amino acids at positions 40-790 has 75% homology to the WD40 domain region of mouse Mapkbp1 protein (gi|47124622) (figure 63a); and the amino acids at positions 40-886 of the AA281370 polypeptide SEQ ID NO: 1564 has 70% homology to rat JNK-binding protein JNKBP1 (gi|34856717) (figure 63b).

Expression of AA281370 transcripts detectable by SEQ ID NO: 1558 (e.g., variant no. 0, 1, 4 and 5, represented in SEQ IDs 1559-1562) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO:1477), was measured similarly. For each RT sample, the expression of SEQ ID NO:1558 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each

RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 58 is a histogram showing over expression of the above-indicated AA281370 transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 6 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figures 58, the expression of AA281370 transcripts detectable by SEQ ID NO:1558 in cancer samples was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 6 fold was found in 8 out of 8 small cell carcinoma, 2 out of 16 squamous cell carcinoma samples, and in 1 out of 4 large cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of AA281370 transcripts detectable by SEQ ID NO:1558 in lung cancer samples versus the normal lung samples was determined by T test as 8.58E-07.

Threshold of 6 fold overexpression was found to differentiate between cancer and normal samples with P value of 4.81E-02 as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

According to the present invention, AA281370 transcripts are a non-limiting example of a marker for diagnosing lung cancer. The AA281370 marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to AA281370 as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for

the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: AA281370-forward primer (SEQ ID NO: 1556): GGTTCGGATGGACTACACTTTGTC; and AA281370-Reverse primer (SEQ ID NO: 1557): CCACGTACTTCTGGGTGATGTC.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: AA281370-amplicon (SEQ ID NO: 1558):
GGTTCGGATGGACTACACTTTGTCCGTACCCACCACGTAGCAGAGAAAACCACCTT
GTATGACATGGACATTGACATCACCCAGAAGTACGTGG.

According to other preferred embodiments of the present invention, AA281370 or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, AA281370 splice variants, as depicted in SEQ ID NO:1558 (e.g., variants no: 0, 1, 4 and 5), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of AA281370 comprises segment_TAA seg 10 SEQ ID NO: 1567, Also optionally and more preferably, any suitable method may be used for detecting a fragment such as AA281370_segment_TAA seg 10 SEQ ID NO: 1567 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments, the present invention also optionally and preferably encompasses AA281370 splice variants containing the unique segments as depicted in SEQ ID NO: 1568, for example transcripts 4 and 5, as depicted in SEQ ID NOs: 1561 and 1562, comprises a biomarker for detecting lung cancer.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to AA281370 as described above, including but not limited to SEQ ID NOs: 1563- 1566. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequence of the proteins SEQ ID NOs: 1563-1566, as depicted in SEQ ID NOs: 1569, 1570 and 1571.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to AA281370 as described above, optionally for any application.

Expression of Sulfatase 1 (SULF1)-[Z21368], transcripts which are detectable by amplicon as depicted in SEQ ID NO:1574 in normal and cancerous lung tissues

SULF1 is a secreted protein which is found in the extracellular matrix. It is known to be downregulated in many epithelial cancer types.

Expression of Sulfatase 1 (SULF1) transcripts detectable by SEQ ID NO:1574 (e.g., variant no. 13 and 14, represented in SEQ ID 1578, 1579) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO:1477), was measured similarly. For each RT sample, the expression of SEQ ID NO: 1574 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 59 is a histogram showing over expression of the above-indicated Sulfatase 1 (SULF1) transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 8 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 59, the expression of Sulfatase 1 (SULF1) transcripts detectable by SEQ ID NO: 1574 in cancer samples originate from non-cell carcinoma was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 8 fold was found in 11 out of 15 adenocarcinoma samples, 11 out of 16 squamous cell carcinoma samples, and in 4 out of 4 large cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of Sulfatase 1 (SULF1) transcripts detectable by SEQ ID NO: 1574 in lung cancer samples versus the normal lung samples was determined by T test as 3.18E-07.

Threshold of 8 fold overexpression was found to differentiate between cancer and normal samples with P value of 1.18E-04 as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

According to the present invention, Sulfatase 1 (SULF1) is a non-limiting example of a marker for diagnosing lung cancer. The Sulfatase 1 (SULF1) marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to Sulfatase 1 (SULF1) as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: Sulfatase 1 (SULF1) - forward primer (SEQ ID NO: 1572): ACTCACTCAGAGACTAACACAAAGGAAG; and Sulfatase 1 (SULF1) - Reverse primer (SEQ ID NO: 1573): AGTATGGGAAGAATTTACTGGTCACA.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: Sulfatase 1 (SULF1) –amplicon (SEQ ID NO: 1574):

ACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCT
ACAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACT.

According to other preferred embodiments of the present invention, Sulfatase 1 (SULF1) or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, Sulfatase 1 (SULF1) splice variants, as depicted in SEQ ID NO:1578, 1579 (e.g., variants no: 13 and 14), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of Sulfatase 1 (SULF1) comprises segment_TAA seg 5 – SEQ ID NO: 1587. Also optionally and more preferably, any suitable method may be used for detecting a fragment such as Sulfatase 1 (SULF1) _segment_ TAA seg 5 – SEQ ID NO: 1587 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments of the present invention, Sulfatase 1 (SULF1) splice variants containing the unique segments as depicted in SEQ ID NOs: 1588-1591, for example transcripts as depicted in SEQ ID NOs: 1575-1577, comprises a biomarker for detecting lung cancer.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to Sulfatase 1 (SULF1) as described above, including but not limited to SEQ ID NOs:1586, 1580, 1582, 1584. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequence of the protein SEQ ID NO: 1580, 1582, 1584, as depicted in SEQ ID NO: 1581, 1583, 1585, respectively.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to Nucleolar protein 4 (NOL4) as described above, optionally for any application.

Expression of SRY (sex determining region Y)-box 2 (SOX2))-[HUMHMGBOX], transcripts which are detectable by the amplicon as depicted in SEQ ID NO:1594 in normal and cancerous lung tissues

Expression of *SOX2* transcripts detectable by SEQ ID NO:1594 (e.g., variant no. 0 represented by SEQ ID 1595) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO: 1477), was measured similarly. For each RT sample, the expression of SEQ ID NO: 1594 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 60 is a histogram showing over expression of the above-indicated SOX2 transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 5 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 60, the expression of SOX2 transcripts detectable by SEQ ID NO: 1594 in cancer samples originate from lung carcinoma was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 5 fold was found in 4 out of 15 adenocarcinoma samples, 10 out of 16 squamous cell carcinoma samples, in 2 out of 4 large cell carcinoma, and in 7 out of 8 small cell carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of SOX2 transcripts detectable by SEQ ID NO: 1594 in lung cancer samples versus the normal lung samples was determined by T test as 4.38E-05.

Threshold of 5 fold overexpression was found to differentiate between cancer and normal samples with P value of 8.09E-04 as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

According to the present invention, SOX2 is a non-limiting example of a marker for diagnosing lung cancer. The SOX2 marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to SOX2 as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: SOX2 -forward primer (SEQ ID NO: 1592): GGCGGCGGCAGGAT; and SOX2 -Reverse primer (SEQ ID NO: 1593): GTCGGGAGCGCAGGG.

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: SOX2 – amplicon (SEQ ID NO: 1594):

GGCGGCGGCAGGATCGGCCAGAGGAGGAGGGAAGCGCTTTTTTTGATCCTGATTCC AGTTTGCCTCTCTCTTTTTTTCCCCCAAATTATTCTTCGCCTGATTTTCCTCGCGGAG CCCTGCGCTCCCGAC.

According to other preferred embodiments of the present invention, SOX2 or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, SOX2 splice variants, as depicted in SEQ ID NO:1595 (e.g., variants no: 0), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of SOX2 comprises segment_TAA seg 2 – SEQ ID NO: 1597. Also optionally and more preferably, any suitable method may be used for detecting a fragment such as SOX2 _segment_ TAA seg 2 – SEQ ID NO: 1597 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to SOX2 as described above, including but not limited to SEQ ID NOs: SEQ ID NO: 1596. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to SOX2 as described above, optionally for any application.

Expression of Plakophilin 1 (ectodermal dysplasia/skin fragility syndrome) (PKP1) -[HSB6PR], transcripts which are detectable by the amplicon as depicted in SEQ ID NO:1600 in normal and cancerous lung tissues

Expression of *PKP1* transcripts detectable by SEQ ID NO:1600 (e.g., variant no. 0, 5 and 6-represented by SEQ IDs 1601-1603) was measured by real time PCR. In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO: 1477), was measured similarly. For each RT sample, the expression of SEQ ID NO: 1600 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel" above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 61 is a histogram showing over expression of the above-indicated PKP1 transcripts in cancerous lung samples relative to the normal samples. The number and

percentage of samples that exhibit at least 7 fold over-expression, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 61, the expression of PKP1 transcripts detectable by SEQ ID NO: 1600 in cancer samples originate from lung carcinoma was significantly higher than in the non-cancerous samples (Sample Nos. 46-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel"). Notably an over-expression of at least 7 fold was found in 11 out of 16 squamous cell carcinoma samples, and in 1 out of 4 large cell carcinoma.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of PKP1 transcripts detectable by SEQ ID NO: 1600 in lung cancer samples versus the normal lung samples was determined by T test as 3.18E-03.

Threshold of 7 fold overexpression was found to differentiate between cancer and normal samples with P value of 3.50E-02 as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

According to the present invention, PKP1 is a non-limiting example of a marker for diagnosing lung cancer. The PKP1 marker of the present invention, can be used alone or in combination, for various uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer. Although optionally any method may be used to detected overexpression and/or differential expression of this marker, preferably a NAT-based technology is used. Therefore, optionally and preferably, any nucleic acid molecule capable of selectively hybridizing to PKP1 as previously defined is also encompassed within the present invention. Primer pairs are also optionally and preferably encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: PKP1 -forward primer (SEQ ID NO: 1598): CCCCAGACTCTGTGCACTTCA; and PKP1 -Reverse primer (SEQ ID NO: 1599): TGGGCTCTGCTCTGTCTTAGTGTA

The present invention also preferably encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: PKP1 –

amplicon (SEQ ID NO: 1600):

CCCCAGACTCTGTGCACTTCAGACCAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTA
TGAGAAAACCTGTGTGGACATCCCTTGGTGTACACTAAGACAGAGCAGAGCCCA

According to other preferred embodiments of the present invention, PKP1 or a fragment thereof comprises a biomarker for detecting lung cancer. Optionally and more preferably, PKP1 splice variants, as depicted in SEQ ID NO: 1601-1603 (e.g., variants no: 0, 5 and 6), or a fragment thereof comprise a biomarker for detecting lung cancer. Optionally and more preferably, the fragment of PKP1 comprises segment_TAA seg 34-SEQ ID NO: 1608. Also optionally and more preferably, any suitable method may be used for detecting a fragment such as PKP1_segment_ TAA seg 34-SEQ ID NO: 1608 for example. Most preferably, NAT-based technology used, such as any nucleic acid molecule capable of specifically hybridizing with the fragment. Optionally and most preferably, a primer pair is used for obtaining the fragment.

According to other preferred embodiments of the present invention, PKP1 splice variants containing the unique segment_8 as depicted in SEQ ID NO: 1607, for example variant 6, as depicted in SEQ ID NO: 1603, are suitable as biomarkers for detecting lung cancer.

According to still other preferred embodiments, the present invention optionally and preferably encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to PKP1 as described above, including but not limited to SEQ ID NOs: 1604-1606. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

The present invention also optionally and preferably encompasses any nucleic acid sequence or fragment thereof, or amino acid sequence or fragment thereof, corresponding to PKP1 as described above, optionally for any application.

Combined expression of 12 sequences (SEQ ID NO: 1480, 1517, 1529, 1532, 1558, 1574, 1594, 1600, 1616, 1619, 1622, 1625) in normal and cancerous lung tissues.

Expression of several transcripts detectable by SEQ ID NOs: 1480, 1517, 1529, 1532, 1558, 1574, 1594, 1600, 1616, 1619, 1622, 1625 was measured by real time PCR (the expression of each SEQ ID was checked separately). In parallel the expression of four housekeeping genes – PBGD (GenBank Accession No. BC019323; amplicon - SEQ ID NO:1471), HPRT1 (GenBank Accession No. NM_000194; amplicon - SEQ ID NO:1468), Ubiquitin (GenBank Accession No. BC000449; amplicon - SEQ ID NO:1474) and SDHA (GenBank Accession No. NM_004168; amplicon - SEQ ID NO: 1477), was measured similarly. For each RT sample, the expression of SEQ ID NOs: 1480, 1517, 1529, 1532, 1558, 1574, 1594, 1600, 1616, 1619, 1622, 1625 was normalized to the geometric mean of the quantities of the housekeeping genes. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal post-mortem (PM) samples (Sample Nos. 47-50, 90-93, 96-99, Table 2, "Tissue samples in testing panel", above), to obtain a value of fold up-regulation for each sample relative to median of the normal PM samples.

Figure 62 is a histogram showing over expression of the above-indicated transcripts in cancerous lung samples relative to the normal samples. The number and percentage of samples that exhibit at least 10 fold over-expression of at least one of the SEQ IDs, out of the total number of samples tested is indicated in the bottom.

As is evident from Figure 62, an over-expression of at least 10 fold in at least one of the SEQ IDs was found in 15 out of 15 adenocarcinoma samples, 15 out of 16 squamous cell carcinoma samples, 4 out of 4 large cell carcinoma samples, and in 8 out of 8 small-cell samples.

Statistical analysis was applied to verify the significance of these results, as described below. Threshold of 10 fold overexpression of at least one of the amplicons as depicted in SEQ ID NOs: 1480, 1517, 1529, 1532, 1558, 1574, 1594, 1600, 1616, 1619, 1622, 1625, was found to differentiate between cancer and normal samples with P value of 2.37E-08 as checked by exact fisher test.

The above values demonstrate statistical significance of the results.

Kits and Diagnostic Assays and Methods

The markers described with regard to any of Examples above can be used alone, in combination with other markers described above, and/or with other entirely different markers, including but not limited to UbcH10 (see US Patent Application Nos: 60/535,904 and 60/572,122; attorney refs: 27080 and 28045, filed on January 13 and May 19 2004, respectively), Troponin (see US Patent Application No: 60/539,129; attorney ref: 26940), Sim2 (see PCT Application No. WO 2004/012847), PE-10 (SP-A), TTF-1, Cytokeratin 5/6, to aid in the diagnosis of lung cancer. All of these applications are hereby incorporated by reference as if fully set forth herein. These markers can be used in combination with other markers for a number of uses, including but not limited to, prognosis, prediction, screening, early diagnosis, therapy selection and treatment monitoring of lung cancer, and also optionally including staging of the disease. Used together, they may provide more information for the diagnostician, increasing the percentage of true positive and true negative diagnoses and decreasing the percentage of false positive or false negative diagnoses, as compared to the results obtained with a single marker alone.

Assays and methods according to the present invention, as described above, include but are not limited to, immunoassays, hybridization assays and NAT-based assays. The combination of the markers of the present invention with other markers described above, and/or with other entirely different markers to aid in the diagnosis of lung cancer could be carried out as a mix of NAT-based assays, immunoassays and hybridization assays. According to preferred embodiments of the present invention, the assays are NAT-based assays, as described for example with regard to the Examples above.

In yet another aspect, the present invention provides kits for aiding a diagnosis of lung cancer, wherein the kits can be used to detect the markers of the present invention. For example, the kits can be used to detect any one or combination of markers described above, which markers are differentially present in samples of a lung cancer patients and normal patients. The kits of the invention have many applications. For example, the kits can be used to differentiate if a subject has a small cell lung cancer, non-small cell lung cancer, adenocarcinoma, bronchoalveolar-alveolar, squamous cell or large cell carcinomas or has a negative diagnosis,

thus aiding a lung cancer diagnosis. In another example, the kits can be used to identify compounds that modulate expression of the markers in in vitro lung cells or in vivo animal models for lung cancer.

In one embodiment, a kit comprises: (a) a substrate comprising an adsorbent thereon, wherein the adsorbent is suitable for binding a marker, and (b) a washing solution or instructions for making a washing solution, wherein the combination of the adsorbent and the washing solution allows detection of the marker as previously described.

Optionally, the kit can further comprise instructions for suitable operational parameters in the form of a label or a separate insert. For example, the kit may have standard instructions informing a consumer/kit user how to wash the probe after a sample of seminal plasma or other tissue sample is contacted on the probe.

In another embodiment, a kit comprises (a) an antibody that specifically binds to a marker; and (b) a detection reagent. Such kits can be prepared from the materials described above.

In either embodiment, the kit may optionally further comprise a standard or control information, and/or a control amount of material, so that the test sample can be compared with the control information standard and/or control amount to determine if the test amount of a marker detected in a sample is a diagnostic amount consistent with a diagnosis of lung cancer.

Therapeutic applications of splice variants of the present invention

Splice variants described herein (including any polynucleotide, oligonucleotide, polypeptide, peptide or fragments thereof) or antibodies that specifically bind thereto may optionally be used for therapeutic applications, for example to treat the diseases described herein with regard to diagnostic applications thereof. A "variant-treatable" disease refers to any disease that is treatable by using a splice variant of any of the therapeutic proteins according to the present invention. "Treatment" also encompasses prevention, amelioration, elimination and control of the disease and/or pathological condition. The diseases for which such variants may be useful therapeutic agents are described in greater detail below for each of the variants. The variants themselves are described by "cluster" or by gene, as these variants are splice variants of known proteins. Therefore, a "cluster-related disease" or a "variant-related disease" refers to a disease that

may be treated by a particular protein, with regard to the description of such diseases below a therapeutic protein variant according to the present invention.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic ligand, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "modulate", as used herein, refers to a change in the activity of at least one receptor mediated activity. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of a ligand.

METHODS OF TREATMENT

As mentioned hereinabove the novel therapeutic protein variants of the present invention and compositions derived therefrom (i.e., peptides, oligonucleotides) can be used to treat cluster-related diseases.

Thus, according to an additional aspect of the present invention there is provided a method of treating cluster-related disease in a subject.

The subject according to the present invention is a mammal, preferably a human which has at least one type of the cluster-related diseases described hereinabove.

As mentioned hereinabove, the biomolecular sequences of the present invention can be used to treat subjects with the above-described diseases.

The subject according to the present invention is a mammal, preferably a human which is diagnosed with one of the diseases described hereinabove, or alternatively is predisposed to having one of the diseases described hereinabove.

As used herein the term “treating” refers to preventing, curing, reversing, attenuating, alleviating, minimizing, suppressing or halting the deleterious effects of the above-described diseases.

Treating, according to the present invention, can be effected by specifically upregulating or alternatively downregulating the expression of at least one of the polypeptides of the present invention in the subject.

Optionally, upregulation may be effected by administering to the subject at least one of the polypeptides of the present invention (e.g., recombinant or synthetic) or an active portion thereof, as described herein. However, since the bioavailability of large polypeptides may potentially be relatively small due to high degradation rate and low penetration rate, administration of polypeptides is preferably confined to small peptide fragments (e.g., about 100 amino acids). The polypeptide or peptide may optionally be administered in a pharmaceutical composition, described in more detail below.

It will be appreciated that treatment of the above-described diseases according to the present invention may be combined with other treatment methods known in the art (i.e., combination therapy). Thus, treatment of malignancies using the agents of the present invention may be combined with, for example, radiation therapy, antibody therapy and/or chemotherapy.

Alternatively or additionally, an upregulating method may optionally be effected by specifically upregulating the amount (optionally expression) in the subject of at least one of the polypeptides of the present invention or active portions thereof.

As is mentioned hereinabove and in the Examples section which follows, the biomolecular sequences of this aspect of the present invention may be used as valuable therapeutic tools in the treatment of diseases in which altered activity or expression of the wild-type gene product is known to contribute to disease onset or progression. For example in case a disease is caused by overexpression of a membrane bound receptor, a soluble variant thereof may be used as an antagonist which competes with the receptor for binding the ligand, to thereby terminate signaling from the receptor.

Examples of such diseases are listed in the Examples section which follows.

It will be appreciated that the polypeptides of the present invention may also have agonistic properties. These include increasing the stability of the ligand (e.g., IL-4), protection from proteolysis and modification of the pharmacokinetic properties of the ligand (i.e., increasing the half-life of the ligand, while decreasing the clearance thereof). As such, the biomolecular sequences of this aspect of the present invention may be used to treat conditions or diseases in

which the wild-type gene product plays a favorable role, for example, increasing angiogenesis in cases of diabetes or ischemia.

Upregulating expression of the therapeutic protein variants of the present invention may be effected via the administration of at least one of the exogenous polynucleotide sequences of the present invention, ligated into a nucleic acid expression construct designed for expression of coding sequences in eukaryotic cells (e.g., mammalian cells), as described above. Accordingly, the exogenous polynucleotide sequence may be a DNA or RNA sequence encoding the variants of the present invention or active portions thereof.

It will be appreciated that the nucleic acid construct can be administered to the individual employing any suitable mode of administration, described hereinbelow (i.e., in-vivo gene therapy). Alternatively, the nucleic acid construct is introduced into a suitable cell via an appropriate gene delivery vehicle/method (transfection, transduction, homologous recombination, etc.) and an expression system as needed and then the modified cells are expanded in culture and returned to the individual (i.e., ex-vivo gene therapy). Nucleic acid constructs are described in greater detail above.

It will be appreciated that the present methodology may also be effected by specifically upregulating the expression of the variants of the present invention endogenously in the subject. Agents for upregulating endogenous expression of specific splice variants of a given gene include antisense oligonucleotides, which are directed at splice sites of interest, thereby altering the splicing pattern of the gene. This approach has been successfully used for shifting the balance of expression of the two isoforms of Bcl-x [Taylor (1999) Nat. Biotechnol. 17:1097-1100; and Mercatante (2001) J. Biol. Chem. 276:16411-16417]; IL-5R [Karras (2000) Mol. Pharmacol. 58:380-387]; and c-myc [Giles (1999) Antisense Acid Drug Dev. 9:213-220].

For example, interleukin 5 and its receptor play a critical role as regulators of hematopoiesis and as mediators in some inflammatory diseases such as allergy and asthma. Two alternatively spliced isoforms are generated from the IL-5R gene, which include (i.e., long form) or exclude (i.e., short form) exon 9. The long form encodes for the intact membrane-bound receptor, while the shorter form encodes for a secreted soluble non-functional receptor. Using 2'-O-MOE-oligonucleotides specific to regions of exon 9, Karras and co-workers (supra) were able to significantly decrease the expression of the wild type receptor and increase the expression of the shorter isoforms. Design and synthesis of oligonucleotides which can be used according to the

present invention are described hereinbelow and by Sazani and Kole (2003) Progress in Moleclular and Subcellular Biology 31:217-239.

Upregulating expression of the polypeptides of the present invention in a subject may be effected via the administration of at least one of the exogenous polynucleotide sequences of the present invention (e.g., SEQ ID NOs: 3, 7, 11, 15, 19, 23, 27, 31, 35, 39 or 43) ligated into a nucleic acid expression construct designed for expression of coding sequences in eukaryotic cells (e.g., mammalian cells). Accordingly, the exogenous polynucleotide sequence may be a DNA or RNA sequence encoding the variants of the present invention or active portions thereof.

It will be appreciated that the nucleic acid construct can be administered to the individual employing any suitable mode of administration, described hereinbelow (i.e., in-vivo gene therapy). Alternatively, the nucleic acid construct is introduced into a suitable cell via an appropriate gene delivery vehicle/method (transfection, transduction, homologous recombination, etc.) and an expression system as needed and then the modified cells are expanded in culture and returned to the individual (i.e., ex-vivo gene therapy).

Preferably, the promoter utilized by the nucleic acid construct of the present invention is active in the specific cell population transformed. Examples of cell type-specific and/or tissue-specific promoters include promoters, such as albumin that is liver specific [Pinkert et al., (1987) Genes Dev. 1:268-277], lymphoid specific promoters [Calame et al., (1988) Adv. Immunol. 43:235-275]; in particular promoters of T-cell receptors [Winoto et al., (1989) EMBO J. 8:729-733] and immunoglobulins; [Banerji et al. (1983) Cell 33729-740], neuron-specific promoters such as the neurofilament promoter [Byrne et al. (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477], pancreas-specific promoters [Edlunch et al. (1985) Science 230:912-916] or mammary gland-specific promoters such as the milk whey promoter (U.S. Pat. No. 4,873,316 and European Patent Application No. EP 264,166).

Examples of suitable constructs include, but are not limited to, pcDNA3, pcDNA3.1 (+/-), pGL3, PzeoSV2 (+/-), pDisplay, pEF/myc/cyto, pCMV/myc/cyto each of which is commercially available from Invitrogen Co. (www.invitrogen.com). Examples of retroviral vector and packaging systems are those sold by Clontech, San Diego, Calif., including Retro-X vectors pLNCX and pLXSN, which permit cloning into multiple cloning sites and the trasgene is transcribed from CMV promoter. Vectors derived from Mo-MuLV are also included such as pBabe, where the transgene will be transcribed from the 5'LTR promoter.

Currently preferred in vivo nucleic acid transfer techniques include transfection with viral or non-viral constructs, such as adenovirus, lentivirus, Herpes simplex I virus, or adeno-associated virus (AAV) and lipid-based systems. Useful lipids for lipid-mediated transfer of the gene are, for example, DOTMA, DOPE, and DC-Chol [Tonkinson et al., Cancer Investigation, 14(1): 54-65 (1996)]. The most preferred constructs for use in gene therapy are viruses, most preferably adenoviruses, AAV, lentiviruses, or retroviruses. A viral construct such as a retroviral construct includes at least one transcriptional promoter/enhancer or locus-defining element(s), or other elements that control gene expression by other means such as alternate splicing, nuclear RNA export, or post-translational modification of messenger. Such vector constructs also include a packaging signal, long terminal repeats (LTRs) or portions thereof, and positive and negative strand primer binding sites appropriate to the virus used, unless it is already present in the viral construct. In addition, such a construct typically includes a signal sequence for secretion of the peptide from a host cell in which it is placed. Preferably the signal sequence for this purpose is a mammalian signal sequence or the signal sequence of the polypeptide variants of the present invention. Optionally, the construct may also include a signal that directs polyadenylation, as well as one or more restriction sites and a translation termination sequence. By way of example, such constructs will typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second-strand DNA synthesis, and a 3' LTR or a portion thereof. Other vectors can be used that are non-viral, such as cationic lipids, polylysine, and dendrimers.

It will be appreciated that the present methodology may also be performed by specifically upregulating the expression of the splice variants of the present invention endogenously in the subject. Agents for upregulating endogenous expression of specific splice variants of a given gene include antisense oligonucleotides, which are directed at splice sites of interest, thereby altering the splicing pattern of the gene. This approach has been successfully used for shifting the balance of expression of the two isoforms of Bcl-x [Taylor (1999) Nat. Biotechnol. 17:1097-1100; and Mercatante (2001) J. Biol. Chem. 276:16411-16417]; IL-5R [Karras (2000) Mol. Pharmacol. 58:380-387]; and c-myc [Giles (1999) Antisense Acid Drug Dev. 9:213-220].

For example, interleukin 5 and its receptor play a critical role as regulators of hematopoiesis and as mediators in some inflammatory diseases such as allergy and asthma. Two alternatively spliced isoforms are generated from the IL-5R gene, which include (i.e., long form) or exclude (i.e., short form) exon 9. The long form encodes for the intact membrane-bound receptor, while the

shorter form encodes for a secreted soluble non-functional receptor. Using 2'-O-MOE-oligonucleotides specific to regions of exon 9, Karras and co-workers (supra) were able to significantly decrease the expression of the wild type receptor and increase the expression of the shorter isoforms. Design and synthesis of oligonucleotides which can be used according to the present invention are described hereinbelow and by Sazani and Kole (2003) Progress in Moleclular and Subcellular Biology 31:217-239.

Treatment can preferably effected by agents which are capable of specifically downregulating expression (or activity) of at least one of the polypeptide variants of the present invention.

Down regulating the expression of the therapeutic protein variants of the present invention may be achieved using oligonucleotide agents such as those described in greater detail below.

SiRNA molecules - Small interfering RNA (siRNA) molecules can be used to down-regulate expression of the therapeutic protein variants of the present invention. RNA interference is a two-step process. The first step, which is termed as the initiation step, input dsRNA is digested into 21-23 nucleotide (nt) small interfering RNAs (siRNA), probably by the action of Dicer, a member of the RNase III family of dsRNA-specific ribonucleases, which processes (cleaves) dsRNA (introduced directly or via a transgene or a virus) in an ATP-dependent manner. Successive cleavage events degrade the RNA to 19-21 bp duplexes (siRNA), each with 2-nucleotide 3' overhangs [Hutvagner and Zamore Curr. Opin. Genetics and Development 12:225-232 (2002); and Bernstein Nature 409:363-366 (2001)].

In the effector step, the siRNA duplexes bind to a nuclease complex to from the RNA-induced silencing complex (RISC). An ATP-dependent unwinding of the siRNA duplex is required for activation of the RISC. The active RISC then targets the homologous transcript by base pairing interactions and cleaves the mRNA into 12 nucleotide fragments from the 3' terminus of the siRNA [Hutvagner and Zamore Curr. Opin. Genetics and Development 12:225-232 (2002); Hammond et al. (2001) Nat. Rev. Gen. 2:110-119 (2001); and Sharp Genes. Dev. 15:485-90 (2001)]. Although the mechanism of cleavage is still to be elucidated, research indicates that each RISC contains a single siRNA and an RNase [Hutvagner and Zamore Curr. Opin. Genetics and Development 12:225-232 (2002)].

Because of the remarkable potency of RNAi, an amplification step within the RNAi pathway has been suggested. Amplification could occur by copying of the input dsRNAs which

would generate more siRNAs, or by replication of the siRNAs formed. Alternatively or additionally, amplification could be effected by multiple turnover events of the RISC [Hammond et al. Nat. Rev. Gen. 2:110-119 (2001), Sharp Genes. Dev. 15:485-90 (2001); Hutvagner and Zamore Curr. Opin. Genetics and Development 12:225-232 (2002)]. For more information on RNAi see the following reviews Tuschl ChemBiochem. 2:239-245 (2001); Cullen Nat. Immunol. 3:597-599 (2002); and Brantl Biochem. Biophys. Act. 1575:15-25 (2002).

Synthesis of RNAi molecules suitable for use with the present invention can be effected as follows. First, the mRNA sequence is scanned downstream of the AUG start codon for AA dinucleotide sequences. Occurrence of each AA and the 3' adjacent 19 nucleotides is recorded as potential siRNA target sites. Preferably, siRNA target sites are selected from the open reading frame, as untranslated regions (UTRs) are richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNA endonuclease complex [Tuschl ChemBiochem. 2:239-245]. It will be appreciated though, that siRNAs directed at untranslated regions may also be effective, as demonstrated for GAPDH wherein siRNA directed at the 5' UTR mediated about 90 % decrease in cellular GAPDH mRNA and completely abolished protein level (www.ambion.com/techlib/tn/91/912.html).

Second, potential target sites are compared to an appropriate genomic database (e.g., human, mouse, rat etc.) using any sequence alignment software, such as the BLAST software available from the NCBI server (www.ncbi.nlm.nih.gov/BLAST/). Putative target sites which exhibit significant homology to other coding sequences are filtered out.

Qualifying target sequences are selected as template for siRNA synthesis. Preferred sequences are those including low G/C content as these have proven to be more effective in mediating gene silencing as compared to those with G/C content higher than 55 %. Several target sites are preferably selected along the length of the target gene for evaluation. Target sites are selected from the unique nucleotide sequences of each of the polynucleotides of the present invention, such that each polynucleotide is specifically down regulated. For better evaluation of the selected siRNAs, a negative control is preferably used in conjunction. Negative control siRNA preferably include the same nucleotide composition as the siRNAs but lack significant homology to the genome. Thus, a scrambled nucleotide sequence of the siRNA is preferably used, provided it does not display any significant homology to any other gene.

DNAzyme molecules - Another agent capable of downregulating expression of the polypeptides of the present invention is a DNAzyme molecule capable of specifically cleaving an mRNA transcript or DNA sequence of the polynucleotides of the present invention. DNAzymes are single-stranded polynucleotides which are capable of cleaving both single and double stranded target sequences (Breaker, R.R. and Joyce, G. Chemistry and Biology 1995;2:655; Santoro, S.W. & Joyce, G.F. Proc. Natl, Acad. Sci. USA 1997;943:4262) A general model (the "10-23" model) for the DNAzyme has been proposed. "10-23" DNAzymes have a catalytic domain of 15 deoxyribonucleotides, flanked by two substrate-recognition domains of seven to nine deoxyribonucleotides each. This type of DNAzyme can effectively cleave its substrate RNA at purine:pyrimidine junctions (Santoro, S.W. & Joyce, G.F. Proc. Natl, Acad. Sci. USA 199; for rev of DNAzymes see Khachigian, LM [Curr Opin Mol Ther 4:119-21 (2002)].

Target sites for DNAzymes are selected from the unique nucleotide sequences of each of the polynucleotides of the present invention, such that each polynucleotide is specifically down regulated.

Examples of construction and amplification of synthetic, engineered DNAzymes recognizing single and double-stranded target cleavage sites have been disclosed in U.S. Pat. No. 6,326,174 to Joyce et al. DNAzymes of similar design directed against the human Urokinase receptor were recently observed to inhibit Urokinase receptor expression, and successfully inhibit colon cancer cell metastasis in vivo (Itoh et al , 20002, Abstract 409, Ann Meeting Am Soc Gen Ther www.asgt.org). In another application, DNAzymes complementary to bcr-ab1 oncogenes were successful in inhibiting the oncogenes expression in leukemia cells, and lessening relapse rates in autologous bone marrow transplant in cases of CML and ALL.

Antisense molecules - Downregulation of the polynucleotides of the present invention can also be effected by using an antisense polynucleotide capable of specifically hybridizing with an mRNA transcript encoding the polypeptide variants of the present invention.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences, which are complementary to a specific DNA or RNA sequence.

The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules also include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form

duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand. Antisense oligonucleotides are also used for modulation of alternative splicing in vivo and for diagnostics in vivo and in vitro (Khelifi C. et al., 2002, Current Pharmaceutical Design 8:451-1466; Sazani, P., and Kole. R. Progress in Molecular and Cellular Biology, 2003, 31:217-239).

Design of antisense molecules which can be used to efficiently downregulate expression of the polypeptides of the present invention must be effected while considering two aspects important to the antisense approach. The first aspect is delivery of the oligonucleotide into the cytoplasm of the appropriate cells, while the second aspect is design of an oligonucleotide which specifically binds the designated mRNA within cells in a way which inhibits translation thereof.

The prior art teaches of a number of delivery strategies which can be used to efficiently deliver oligonucleotides into a wide variety of cell types [see, for example, Luft J Mol Med 76: 75-6 (1998); Kronenwett et al. Blood 91: 852-62 (1998); Rajur et al. Bioconjug Chem 8: 935-40 (1997); Lavigne et al. Biochem Biophys Res Commun 237: 566-71 (1997) and Aoki et al. (1997) Biochem Biophys Res Commun 231: 540-5 (1997)].

In addition, algorithms for identifying those sequences with the highest predicted binding affinity for their target mRNA based on a thermodynamic cycle that accounts for the energetics of structural alterations in both the target mRNA and the oligonucleotide are also available [see, for example, Walton et al. Biotechnol Bioeng 65: 1-9 (1999)].

Such algorithms have been successfully used to implement an antisense approach in cells. For example, the algorithm developed by Walton et al. enabled scientists to successfully design antisense oligonucleotides for rabbit beta-globin (RBG) and mouse tumor necrosis factor-alpha (TNF alpha) transcripts. The same research group has more recently reported that the antisense activity of rationally selected oligonucleotides against three model target mRNAs (human lactate dehydrogenase A and B and rat gp130) in cell culture as evaluated by a kinetic PCR technique proved effective in almost all cases, including tests against three different targets in two cell types with phosphodiester and phosphorothioate oligonucleotide chemistries.

In addition, several approaches for designing and predicting efficiency of specific oligonucleotides using an in vitro system were also published (Matveeva et al., Nature Biotechnology 16: 1374 - 1375 (1998)].

Several clinical trials have demonstrated safety, feasibility and activity of antisense oligonucleotides. For example, antisense oligonucleotides suitable for the treatment of cancer have been successfully used [Holmund et al., Curr Opin Mol Ther 1:372-85 (1999)], while treatment of hematological malignancies via antisense oligonucleotides targeting c-myb gene, p53 and Bcl-2 had entered clinical trials and had been shown to be tolerated by patients [Gerwitz Curr Opin Mol Ther 1:297-306 (1999)].

More recently, antisense-mediated suppression of human heparanase gene expression has been reported to inhibit pleural dissemination of human cancer cells in a mouse model [Uno et al., Cancer Res 61:7855-60 (2001)].

Thus, the current consensus is that recent developments in the field of antisense technology which, as described above, have led to the generation of highly accurate antisense design algorithms and a wide variety of oligonucleotide delivery systems, enable an ordinarily skilled artisan to design and implement antisense approaches suitable for downregulating expression of known sequences without having to resort to undue trial and error experimentation.

Target sites for antisense molecules are selected from the unique nucleotide sequences of each of the polynucleotides of the present invention, such that each polynucleotide is specifically down regulated.

Ribozymes - Another agent capable of downregulating expression of the polypeptides of the present invention is a ribozyme molecule capable of specifically cleaving an mRNA transcript encoding the polypeptide variants of the present invention. Ribozymes are being increasingly used for the sequence-specific inhibition of gene expression by the cleavage of mRNAs encoding proteins of interest [Welch et al., Curr Opin Biotechnol. 9:486-96 (1998)]. The possibility of designing ribozymes to cleave any specific target RNA has rendered them valuable tools in both basic research and therapeutic applications. In therapeutics area, ribozymes have been exploited to target viral RNAs in infectious diseases, dominant oncogenes in cancers and specific somatic mutations in genetic disorders [Welch et al., Clin Diagn Virol. 10:163-71 (1998)]. Most notably, several ribozyme gene therapy protocols for HIV patients are already in Phase 1 trials. More recently, ribozymes have been used for transgenic animal research, gene target validation and pathway elucidation. Several ribozymes are in various stages of clinical trials. ANGIOZYME was the first chemically synthesized ribozyme to be studied in human clinical trials. ANGIOZYME specifically inhibits formation of the VEGF-r (Vascular Endothelial Growth Factor receptor), a key

component in the angiogenesis pathway. Ribozyme Pharmaceuticals, Inc., as well as other firms have demonstrated the importance of anti-angiogenesis therapeutics in animal models. HEPTAZYME, a ribozyme designed to selectively destroy Hepatitis C Virus (HCV) RNA, was found effective in decreasing Hepatitis C viral RNA in cell culture assays (Ribozyme Pharmaceuticals, Incorporated - WEB home page).

Alternatively, down regulation of the polypeptide variants of the present invention may be achieved at the polypeptide level using downregulating agents such as antibodies or antibody fragments capabale of specifically binding the polypeptides of the present invention and inhibiting the activity thereof (i.e., neutralizing antibodies). Such antibodies can be directed for example, to the heterodimerizing domain on the variant, or to a putative ligand binding domain. Further description of antibodies and methods of generating same is provided below.

PHARMACEUTICAL COMPOSITIONS AND DELIVERY THEREOF

The present invention features a pharmaceutical composition comprising a therapeutically effective amount of a therapeutic agent according to the present invention, which is preferably a therapeutic protein variant as described herein. Optionally and alternatively, the therapeutic agent could be an antibody or an oligonucleotide that specifically recognizes and binds to the therapeutic protein variant, but not to the corresponding full length known protein.

Alternatively, the pharmaceutical composition of the present invention includes a therapeutically effective amount of at least an active portion of a therapeutic protein variant polypeptide.

The pharmaceutical composition according to the present invention is preferably used for the treatment of cluster-related diseases.

"Treatment" refers to both therapeutic treatment and prophylactic or preventative measures. Those in need of treatment include those already with the disorder as well as those in which the disorder is to be prevented. Hence, the mammal to be treated herein may have been diagnosed as having the disorder or may be predisposed or susceptible to the disorder. "Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, domestic and farm animals, and zoo, sports, or pet animals, such as dogs, horses, cats, cows, etc. Preferably, the mammal is human.

A "disorder" is any condition that would benefit from treatment with the agent according to the present invention. This includes chronic and acute disorders or diseases including those pathological conditions which predispose the mammal to the disorder in question. Non-limiting examples of disorders to be treated herein are described with regard to specific examples given herein.

The term "therapeutically effective amount" refers to an amount of agent according to the present invention that is effective to treat a disease or disorder in a mammal. In the case of cancer, the therapeutically effective amount of the agent may reduce the number of cancer cells; reduce the tumor size; inhibit (i.e., slow to some extent and preferably stop) cancer cell infiltration into peripheral organs; inhibit (i.e., slow to some extent and preferably stop) tumor metastasis; inhibit, to some extent, tumor growth; and/or relieve to some extent one or more of the symptoms associated with the cancer. To the extent the agent may prevent growth and/or kill existing cancer cells, it may be cytostatic and/or cytotoxic. For cancer therapy, efficacy can, for example, be measured by assessing the time to disease progression (TTP) and/or determining the response rate (RR).

The therapeutic agents of the present invention can be provided to the subject per se, or as part of a pharmaceutical composition where they are mixed with a pharmaceutically acceptable carrier.

As used herein a "pharmaceutical composition" refers to a preparation of one or more of the active ingredients described herein with other chemical components such as physiologically suitable carriers and excipients. The purpose of a pharmaceutical composition is to facilitate administration of a compound to an organism.

Herein the term "active ingredient" refers to the preparation accountable for the biological effect.

Hereinafter, the phrases "physiologically acceptable carrier" and "pharmaceutically acceptable carrier" which may be interchangeably used refer to a carrier or a diluent that does not cause significant irritation to an organism and does not abrogate the biological activity and properties of the administered compound. An adjuvant is included under these phrases. One of the ingredients included in the pharmaceutically acceptable carrier can be for example polyethylene glycol (PEG), a biocompatible polymer with a wide range of solubility in both organic and aqueous media (Mutter et al. (1979).

Herein the term "excipient" refers to an inert substance added to a pharmaceutical composition to further facilitate administration of an active ingredient. Examples, without limitation, of excipients include calcium carbonate, calcium phosphate, various sugars and types of starch, cellulose derivatives, gelatin, vegetable oils and polyethylene glycols.

Techniques for formulation and administration of drugs may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition, which is incorporated herein by reference.

Suitable routes of administration may, for example, include oral, rectal, transmucosal, especially transnasal, intestinal or parenteral delivery, including intramuscular, subcutaneous and intramedullary injections as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Alternately, one may administer a preparation in a local rather than systemic manner, for example, via injection of the preparation directly into a specific region of a patient's body.

Pharmaceutical compositions of the present invention may be manufactured by processes well known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes.

Pharmaceutical compositions for use in accordance with the present invention may be formulated in conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active ingredients into preparations which, can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

For injection, the active ingredients of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hank's solution, Ringer's solution, or physiological salt buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for oral ingestion by a patient. Pharmacological preparations for oral use can be made using a solid excipient, optionally grinding the resulting

mixture, and processing the mixture of granules, after adding suitable auxiliaries if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carbomethylcellulose; and/or physiologically acceptable polymers such as polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, titanium dioxide, lacquer solutions and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical compositions, which can be used orally, include push-fit capsules made of gelatin as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules may contain the active ingredients in admixture with filler such as lactose, binders such as starches, lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active ingredients may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for the chosen route of administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by nasal inhalation, the active ingredients for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from a pressurized pack or a nebulizer with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichloro-tetrafluoroethane or carbon dioxide. In the case of a pressurized aerosol, the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in a dispenser may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The preparations described herein may be formulated for parenteral administration, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage

form, e.g., in ampoules or in multidose containers with optionally, an added preservative. The compositions may be suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical compositions for parenteral administration include aqueous solutions of the active preparation in water-soluble form. Additionally, suspensions of the active ingredients may be prepared as appropriate oily or water based injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acids esters such as ethyl oleate, triglycerides or liposomes. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the active ingredients to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, e.g., sterile, pyrogen-free water based solution, before use.

The preparation of the present invention may also be formulated in rectal compositions such as suppositories or retention enemas, using, e.g., conventional suppository bases such as cocoa butter or other glycerides.

Pharmaceutical compositions suitable for use in context of the present invention include compositions wherein the active ingredients are contained in an amount effective to achieve the intended purpose. More specifically, a therapeutically effective amount means an amount of active ingredients effective to prevent, alleviate or ameliorate symptoms of disease or prolong the survival of the subject being treated.

Determination of a therapeutically effective amount is well within the capability of those skilled in the art.

For any preparation used in the methods of the invention, the therapeutically effective amount or dose can be estimated initially from in vitro assays. For example, a dose can be formulated in animal models and such information can be used to more accurately determine useful doses in humans.

Toxicity and therapeutic efficacy of the active ingredients described herein can be determined by standard pharmaceutical procedures in vitro, in cell cultures or experimental animals. The data obtained from these in vitro and cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage may vary depending upon the

dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (See e.g., Fingl, et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1).

Depending on the severity and responsiveness of the condition to be treated, dosing can be of a single or a plurality of administrations, with course of treatment lasting from several days to several weeks or until cure is effected or diminution of the disease state is achieved.

The amount of a composition to be administered will, of course, be dependent on the subject being treated, the severity of the affliction, the manner of administration, the judgment of the prescribing physician, etc.

Compositions including the preparation of the present invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

Pharmaceutical compositions of the present invention may, if desired, be presented in a pack or dispenser device, such as an FDA approved kit, which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. The pack or dispenser may also be accommodated by a notice associated with the container in a form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals, which notice is reflective of approval by the agency of the form of the compositions or human or veterinary administration. Such notice, for example, may be of labeling approved by the U.S. Food and Drug Administration for prescription drugs or of an approved product insert.

IMMUNOGENIC COMPOSITIONS

A therapeutic agent according to the present invention may optionally be a molecule, which promotes a specific immunogenic response against at least one of the polypeptides of the present invention in the subject. The molecule can be polypeptide variants of the present invention, a fragment derived therefrom or a nucleic acid sequence encoding thereof. Although such a molecule can be provided to the subject per se, the agent is preferably administered with an immunostimulant in an immunogenic composiiton. An immunostimulant may be any substance that enhances or potentiates an immune response (antibody and/or cell-mediated) to an exogenous

antigen. Examples of immunostimulants include adjuvants, biodegradable microspheres (e.g., polylactic galactide) and liposomes into which the compound is incorporated (see e.g., U.S. Pat. No. 4,235,877). Vaccine preparation is generally described in, for example, M. F. Powell and M. J. Newman, eds., "Vaccine Design (the subunit and adjuvant approach)," Plenum Press (NY, 1995).

Illustrative immunogenic compositions may contain DNA encoding one or more of the polypeptides as described above, such that the polypeptide is generated in situ. The DNA may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems (see below), bacteria and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, Crit. Rev. Therap. Drug Carrier Systems 15:143-198, 1998, and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the subject (such as a suitable promoter and terminating signal). Bacterial delivery systems involve the administration of a bacterium (such as Bacillus-Calmette-Guerrin) that expresses an immunogenic portion of the polypeptide on its cell surface or secretes such an epitope. In a preferred embodiment, the DNA may be introduced using a viral expression system (e.g., vaccinia or other pox virus, retrovirus, or adenovirus), which may involve the use of a non-pathogenic (defective), replication competent virus. Suitable systems are disclosed, for example, in Fisher-Hoch et al., Proc. Natl. Acad. Sci. USA 86:317-321, 1989; Flexner et al., Ann. N.Y Acad. Sci. 569:86-103, 1989; Flexner et al., Vaccine 8:17-21, 1990; U.S. Pat. Nos. 4,603,112, 4,769,330, and 5,017,487; WO 89/01973; U.S. Pat. No. 4,777,127; GB 2,200,651; EP 0,345,242; WO 91/02805; Berkner, Biotechniques 6:616-627, 1988; Rosenfeld et al., Science 252:431-434, 1991; Kolls et al., Proc. Natl. Acad. Sci. USA 91:215-219, 1994; Kass-Eisler et al., Proc. Natl. Acad. Sci. USA 90:11498-11502, 1993; Guzman et al., Circulation 88:2838-2848, 1993; and Guzman et al., Cir. Res. 73:1202-1207, 1993. Techniques for incorporating DNA into such expression systems are well known to those of ordinary skill in the art. The DNA may also be "naked," as described, for example, in Ulmer et al., Science 259:1745-1749, 1993 and reviewed by Cohen, Science 259:1691-1692, 1993. The uptake of naked DNA may be increased by coating the DNA onto biodegradable beads, which are efficiently transported into the cells.

It will be appreciated that an immunogenic composition may comprise both a polynucleotide and a polypeptide component. Such immunogenic compositions may provide for an enhanced immune response.

Any of a variety of immunostimulants may be employed in the immunogenic compositions of this invention. For example, an adjuvant may be included. Most adjuvants contain a substance designed to protect the antigen from rapid catabolism, such as aluminum hydroxide or mineral oil, and a stimulator of immune responses, such as lipid A, Bortadella pertussis or Mycobacterium tuberculosis derived proteins. Suitable adjuvants are commercially available as, for example, Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, Mich.); Merck Adjuvant 65 (Merck and Company, Inc., Rahway, N.J.); AS-2 (SmithKline Beecham, Philadelphia, Pa.); aluminum salts such as aluminum hydroxide gel (alum) or aluminum phosphate; salts of calcium, iron or zinc; an insoluble suspension of acylated tyrosine; acylated sugars; cationically or anionically derivatized polysaccharides; polyphosphazenes; biodegradable microspheres; monophosphoryl lipid A and quil A. Cytokines, such as GM-CSF or interleukin-2,-7, or -12, may also be used as adjuvants.

The adjuvant composition may be designed to induce an immune response predominantly of the Th1 type. High levels of Th1-type cytokines (e.g., IFN-.gamma., TNF.alpha., IL-2 and IL-12) tend to favor the induction of cell mediated immune responses to an administered antigen. In contrast, high levels of Th2-type cytokines (e.g., IL-4, IL-5, IL-6 and IL-10) tend to favor the induction of humoral immune responses. Following application of an immunogenic composition as provided herein, the subject will support an immune response that includes Th1- and Th2-type responses. The levels of these cytokines may be readily assessed using standard assays. For a review of the families of cytokines, see Mosmann and Coffinan, Ann. Rev. Immunol. 7:145-173, 1989.

Preferred adjuvants for use in eliciting a predominantly Th1-type response include, for example, a combination of monophosphoryl lipid A, preferably 3-de-O-acylated monophosphoryl lipid A (3D-MPL), together with an aluminum salt. MPL adjuvants are available from Corixa Corporation (Seattle, Wash.; see U.S. Pat. Nos. 4,436,727; 4,877,611; 4,866,034 and 4,912,094). CpG-containing oligonucleotides (in which the CpG dinucleotide is unmethylated) also induce a predominantly Thl response. Such oligonucleotides are well known and are described, for example, in WO 96/02555, WO 99/33488 and U.S. Pat. Nos. 6,008,200 and 5,856,462. Immunostimulatory DNA sequences are also described, for example, by Sato et al., Science 273:352, 1996. Another preferred adjuvant is a saponin, preferably QS21 (Aquila Biopharmaceuticals Inc., Framingham, Mass.), which may be used alone or in combination with other adjuvants. For example, an

enhanced system involves the combination of a monophosphoryl lipid A and saponin derivative, such as the combination of QS21 and 3D-MPL as described in WO 94/00153, or a less reactogenic composition where the QS21 is quenched with cholesterol, as described in WO 96/33739. Other preferred formulations comprise an oil-in-water emulsion and tocopherol. A particularly potent adjuvant formulation involving QS21, 3D-MPL and tocopherol in an oil-in-water emulsion is described in WO 95/17210.

Other preferred adjuvants include Montanide ISA 720 (Seppic, France), SAF (Chiron, Calif., United States), ISCOMS (CSL), MF-59 (Chiron), the SBAS series of adjuvants (e.g., SBAS-2 or SBAS-4, available from SmithKline Beecham, Rixensart, Belgium), Detox (Corixa, Hamilton, Mont.), RC-529 (Corixa, Hamilton, Mont.) and other aminoalkyl glucosaminide 4-phosphates (AGPs), such as those described in pending U.S. patent application Ser. Nos. 08/853,826 and 09/074,720.

A delivery vehicle may be employed within the immunogenic composition of the present invention to facilitate production of an antigen-specific immune response that targets tumor cells. Delivery vehicles include antigen presenting cells (APCs), such as dendritic cells, macrophages, B cells, monocytes and other cells that may be engineered to be efficient APCs. Such cells may be genetically modified to increase the capacity for presenting the antigen, to improve activation and/or maintenance of the T cell response, to have anti-tumor effects per se and/or to be immunologically compatible with the receiver (i.e., matched HLA haplotype). APCs may generally be isolated from any of a variety of biological fluids and organs, including tumor and peritumoral tissues, and may be autologous, allogeneic, syngeneic or xenogeneic cells.

Dendritic cells are highly potent APCs (Banchereau and Steinman, Nature 392:245-251, 1998) and have been shown to be effective as a physiological adjuvant for eliciting prophylactic or therapeutic antitumor immunity (see Timmernan and Levy, Ann. Rev. Med. 50:507-529, 1999). In general, dendritic cells may be identified based on their typical shape (stellate in situ, with marked cytoplasmic processes (dendrites) visible in vitro), their ability to take up, process and present antigens with high efficiency and their ability to activate naive T cell responses. Dendritic cells may, of course, be engineered to express specific cell-surface receptors or ligands that are not commonly found on dendritic cells in vivo or ex vivo, and such modified dendritic cells are contemplated by the present invention. As an alternative to dendritic cells, secreted vesicles

antigen-loaded dendritic cells (called exosomes) may be used within an immunogenic composition (see Zitvogel et al., Nature Med. 4:594-600, 1998).

Dendritic cells and progenitors may be obtained from peripheral blood, bone marrow, tumor-infiltrating cells, peritumoral tissues-infiltrating cells, lymph nodes, spleen, skin, umbilical cord blood or any other suitable tissue or fluid. For example, dendritic cells may be differentiated ex vivo by adding a combination of cytokines such as GM-CSF, IL-4, IL-13 and/or TNF.alpha. to cultures of monocytes harvested from peripheral blood. Alternatively, CD34 positive cells harvested from peripheral blood, umbilical cord blood or bone marrow may be differentiated into dendritic cells by adding to the culture medium combinations of GM-CSF, IL-3, TNF.alpha., CD40 ligand, LPS, flt3 ligand and/or other compound(s) that induce differentiation, maturation and proliferation of dendritic cells.

Dendritic cells are categorized as "immature" and "mature" cells, which allows a simple way to discriminate between two well characterized phenotypes. Immature dendritic cells are characterized as APC with a high capacity for antigen uptake and processing, which correlates with the high expression of Fcy receptor and mannose receptor. The mature phenotype is typically characterized by a lower expression of these markers, but a high expression of cell surface molecules responsible for T cell activation such as class I and class II MHC, adhesion molecules (e.g., CD54 and CD11) and costimulatory molecules (e.g., CD40, CD80, CD86 and 4-1BB).

APCs may generally be transfected with at least one polynucleotide encoding a polypeptide of the present invention, such that variant II, or an immunogenic portion thereof, is expressed on the cell surface. Such transfection may take place ex vivo, and a composition comprising such transfected cells may then be used for therapeutic purposes, as described herein. Alternatively, a gene delivery vehicle that targets a dendritic or other antigen presenting cell may be administered to the subject, resulting in transfection that occurs in vivo. In vivo and ex vivo transfection of dendritic cells, for example, may generally be performed using any methods known in the art, such as those described in WO 97/24447, or the gene gun approach described by Mahvi et al., Immunology and cell Biology 75:456-460, 1997. Antigen loading of dendritic cells may be achieved by incubating dendritic cells or progenitor cells with a polypeptide of the present inventio, DNA (naked or within a plasmid vector) or RNA; or with antigen-expressing recombinant bacterium or viruses (e.g., vaccinia, fowlpox, adenovirus or lentivirus vectors). Prior to loading, the polypeptide may be covalently conjugated to an immunological partner that provides T cell

help (e.g., a carrier molecule) such as described above. Alternatively, a dendritic cell may be pulsed with a non-conjugated immunological partner, separately or in the presence of the polypeptide.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a polynucleotide having a sequence of R11723_PEA_1_T5.

2. The isolated polynucleotide of claim 1, comprising a node having a sequence of: R11723_PEA_1_node_13.

3. An isolated polypeptide comprising a polypeptide having a sequence of: R11723_PEA_1_P13.

4. The isolated of claim 3, comprising a chimeric polypeptide encoding for R11723_PEA_1_P13, comprising a first amino acid sequence being at least 95 % homologous to MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEV MEQSA corresponding to amino acids 1 - 63 of Q96AC2, which also corresponds to amino acids 1 - 63 of R11723_PEA_1_P13, and a second amino acid sequence being at least about 95% homologous to a polypeptide having the sequence DTKRTNTLLFEMRHFAKQLTT corresponding to amino acids 64 - 84 of R11723_PEA_1_P13, wherein said first and second amino acid sequences are contiguous and in a sequential order.

4. The isolated polypeptide of claim 4, comprising a tail of R11723_PEA_1_P13, comprising a polypeptide being at least about 95% homologous to the sequence DTKRTNTLLFEMRHFAKQLTT in R11723_PEA_1_P13.

5. The isolated oligonucleotide of claim 1, comprising an amplicon according to SEQ ID NO: 1684.

6. A primer pair, comprising a pair of isolated oligonucleotides capable of amplifying said amplicon of claim 5.

7. The primer pair of claim 6, comprising a pair of isolated oligonucleotides: SEQ NOs 1682 and 1683.

8. An antibody capable of specifically binding to an epitope of an amino acid sequence of claim 3.

9. The antibody of claim 8, wherein said amino acid sequence comprises said tail of claim 4.

10. The antibody of claim 8, wherein said antibody is capable of differentiating between a splice variant having said epitope and a corresponding known protein PSEC.

11. A kit for detecting lung cancer, comprising a kit detecting overexpression of a splice variant according to claim 1.

12. The kit of claim 11, wherein said kit comprises a NAT-based technology.

13. The kit of claim 11, wherein said kit further comprises at least one primer pair capable of selectively hybridizing to a nucleic acid sequence according to claim 1.

14. The kit of claim 11, wherein said kit further comprises at least one oligonucleotide capable of selectively hybridizing to a nucleic acid sequence according to claim 1.

12. A kit for detecting lung cancer, comprising a kit detecting overexpression of a splice variant according to claim 3, said kit comprising an antibody according claim 8.

13. The kit of claim 12, wherein said kit further comprises at least one reagent for performing an ELISA or a Western blot.

14. A method for detecting lung cancer, comprising detecting overexpression of a splice variant according to claim 1.

15. The method of claim 14, wherein said detecting overexpression is performed with a NAT-based technology.

16. A method for detecting lung cancer, comprising detecting overexpression of a splice variant according to claim 3, wherein said detecting overexpression is performed with an immunoassay.

17. The method of claim 16, wherein said immunoassay comprises an antibody according to claim 8.

18. A biomarker capable of detecting lung cancer, comprising a nucleic acid sequence according to claim 1 or a fragment thereof, or an amino acid sequence according to claim 3 or a fragment thereof.

19. A method for screening for lung cancer, comprising detecting lung cancer cells with a biomarker according to claim 18.

20. A method for diagnosing lung cancer, comprising detecting lung cancer cells with a biomarker according to claim 18.

21. A method for monitoring disease progression and/or treatment efficacy and/or relapse of lung cancer, comprising detecting lung cancer cells with a biomarker according to claim 18.

22. A method of selecting a therapy for lung cancer, comprising detecting lung cancer cells with a biomarker according to claim 18 and selecting a therapy according to said detection.

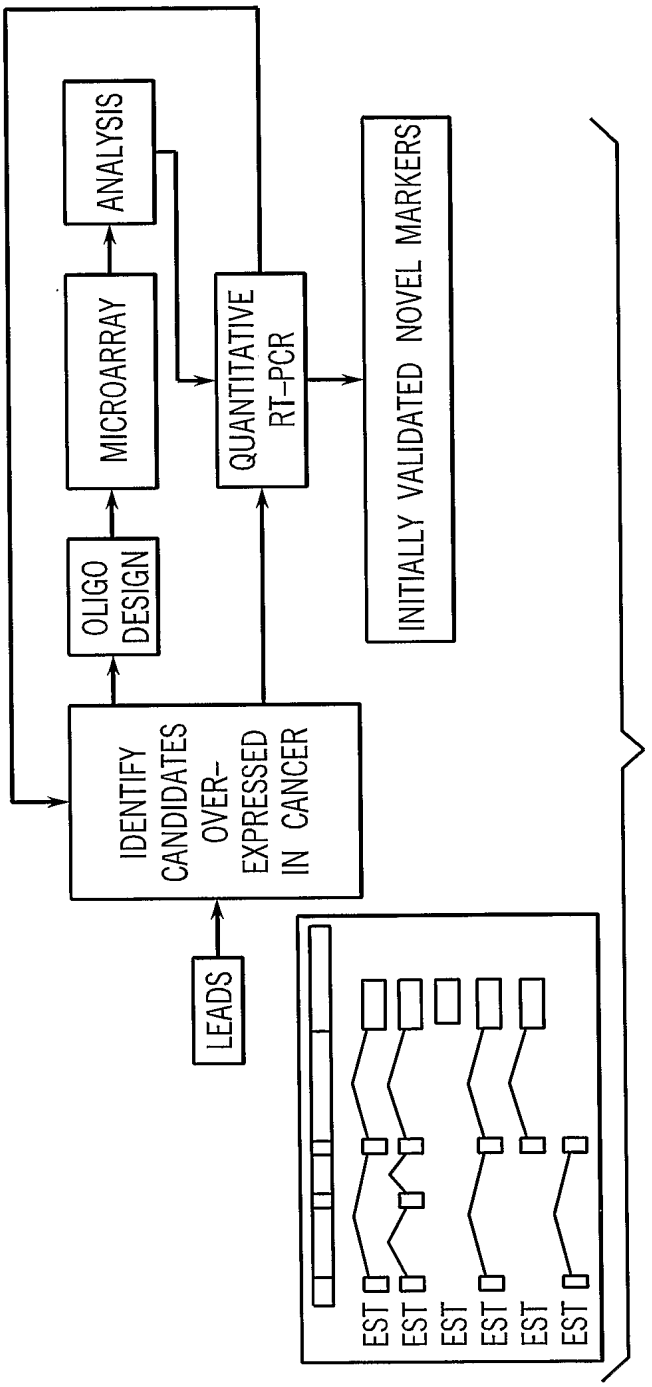

LEADS
IDENTIFY CANDIDATES OVER-EXPRESSED IN CANCER
OLIGO DESIGN
MICROARRAY
ANALYSIS
QUANTITATIVE RT-PCR
INITIALLY VALIDATED NOVEL MARKERS
EST
EST
EST
EST
EST
EST

FIG. 1

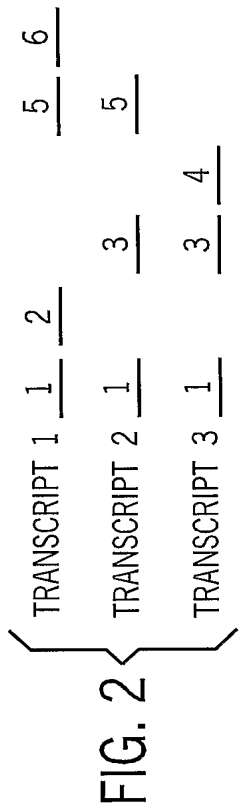

TRANSCRIPT 1
1
2
5
6
TRANSCRIPT 2
1
3
5
TRANSCRIPT 3
1
3
4

FIG. 2

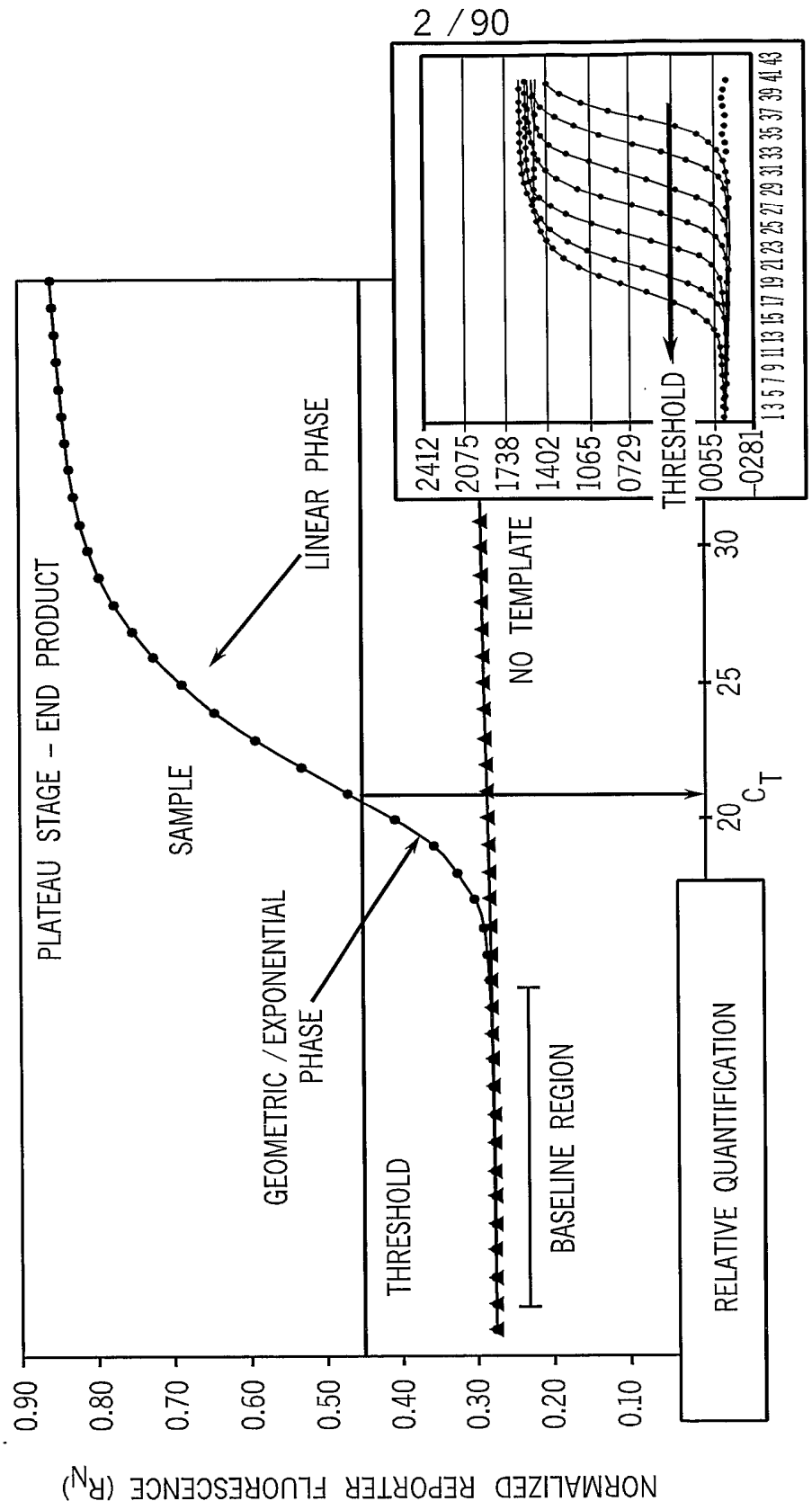

2 / 90
2412
2075
1738
1402
1065
0729
0055
0281
THRESHOLD
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43
PLATEAU STAGE - END PRODUCT
SAMPLE
LINEAR PHASE
GEOMETRIC / EXPONENTIAL PHASE
NO TEMPLATE
THRESHOLD
BASELINE REGION
RELATIVE QUANTIFICATION
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
20
25
30
C_T
NORMALIZED REPORTER FLUORESCENCE (R_N)

FIG. 3

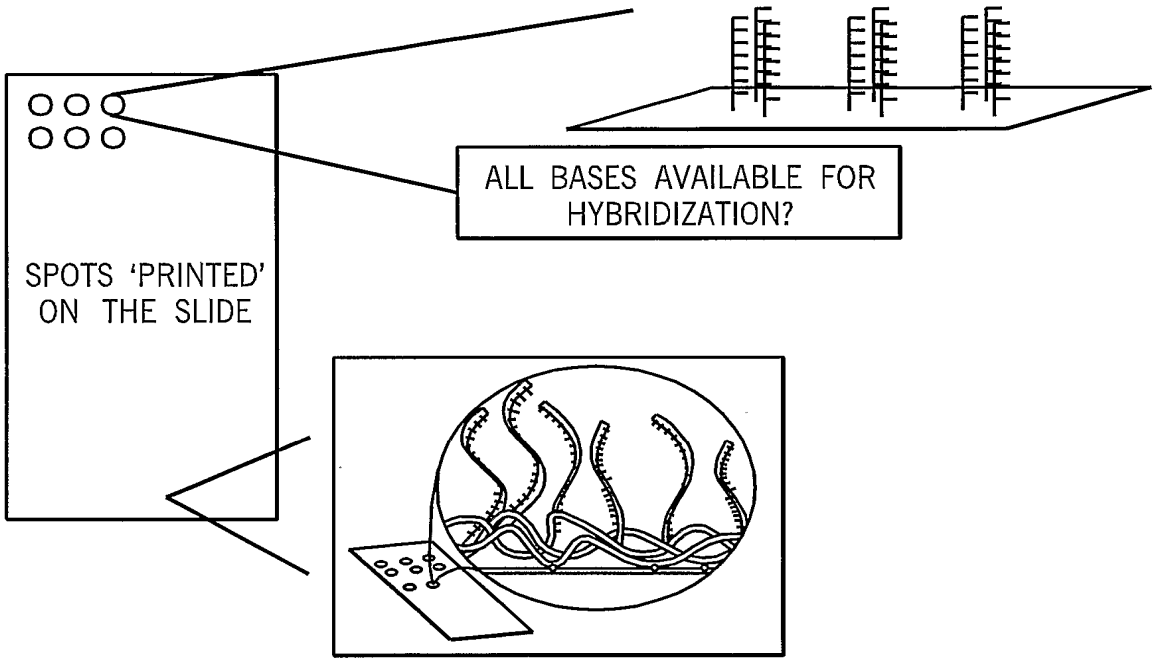

ALL BASES AVAILABLE FOR HYBRIDIZATION?
SPOTS 'PRINTED' ON THE SLIDE

FIG. 4

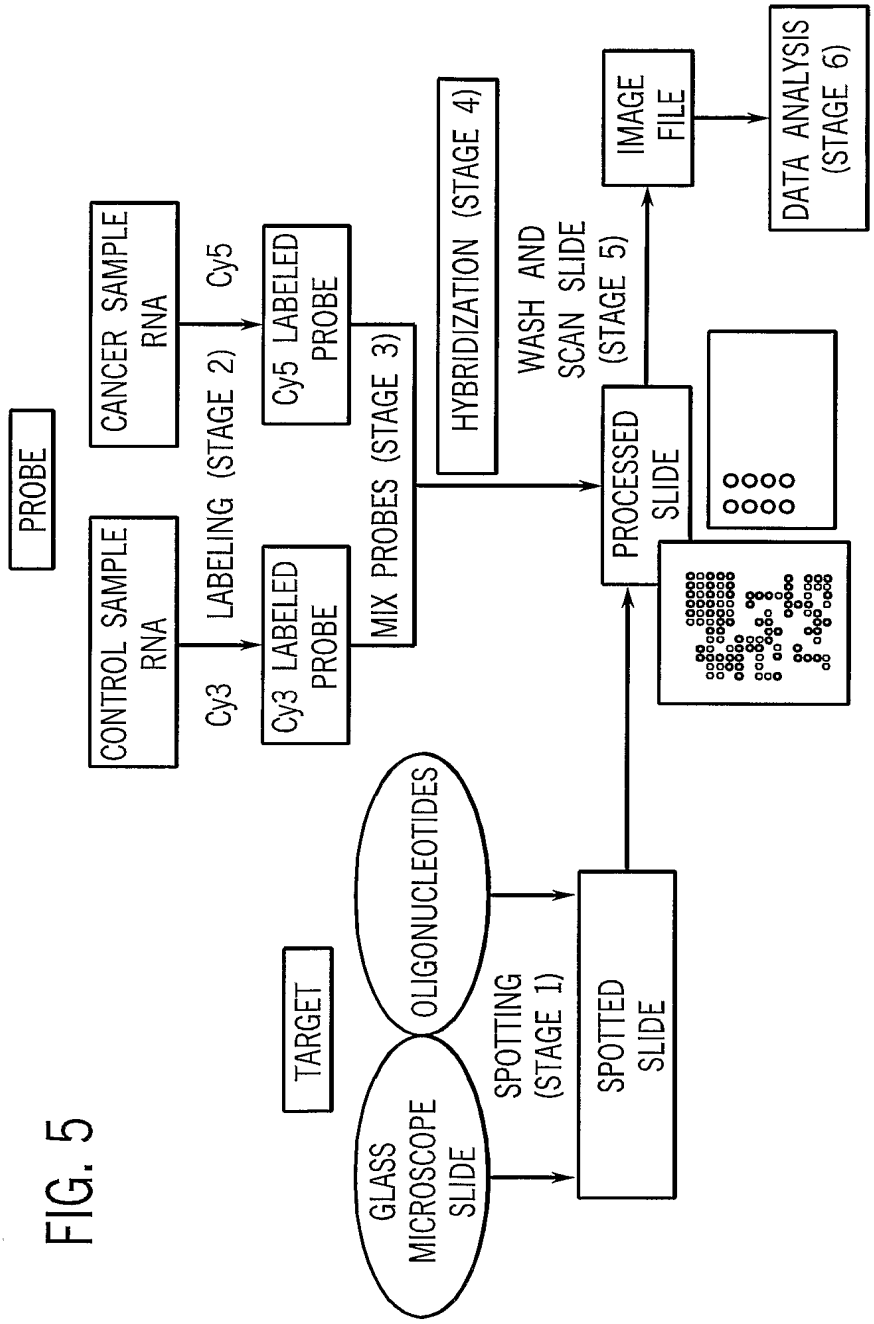

PROBE
CONTROL SAMPLE RNA
Cy3
LABELING (STAGE 2)
Cy3 LABELED PROBE
CANCER SAMPLE RNA
Cy5
Cy5 LABELED PROBE
MIX PROBES (STAGE 3)
HYBRIDIZATION (STAGE 4)
WASH AND SCAN SLIDE (STAGE 5)
PROCESSED SLIDE
IMAGE FILE
DATA ANALYSIS (STAGE 6)
TARGET
OLIGONUCLEOTIDES
GLASS MICROSCOPE SLIDE
SPOTTING (STAGE 1)
SPOTTED SLIDE

FIG. 5

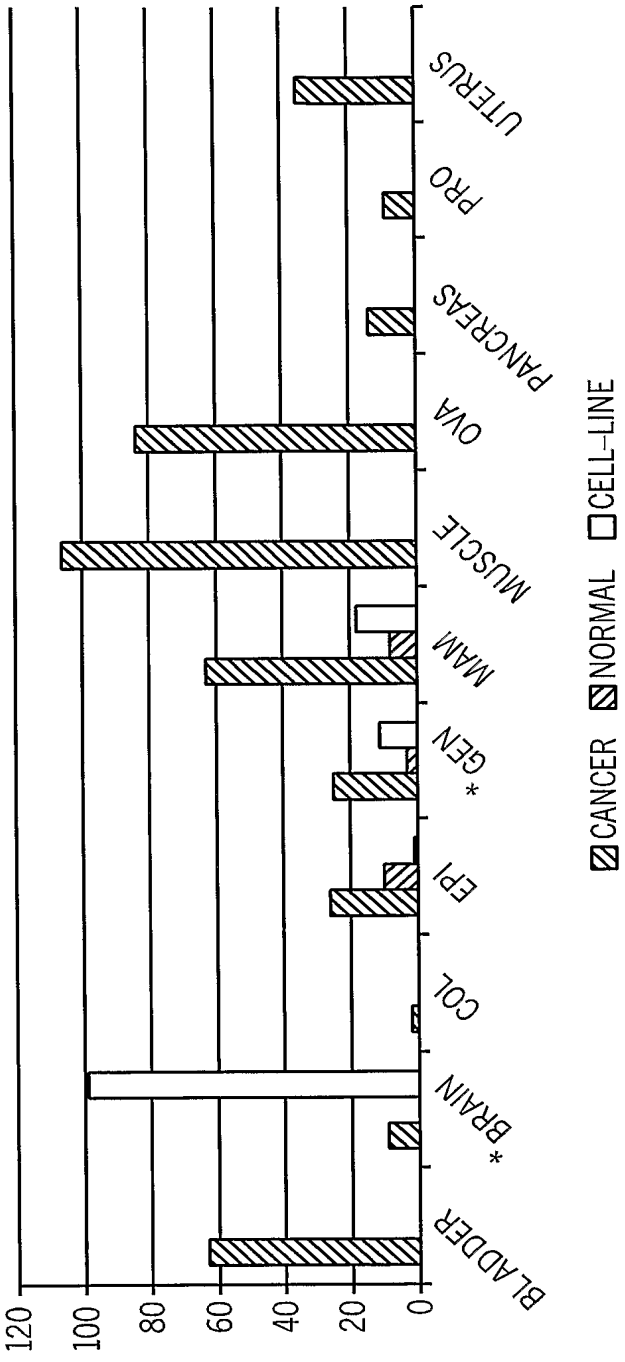

120
100
80
60
40
20
0
BLADDER
* BRAIN
COL
EPI
* GEN
MAM
MUSCLE
OVA
PANCREAS
PRO
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 6

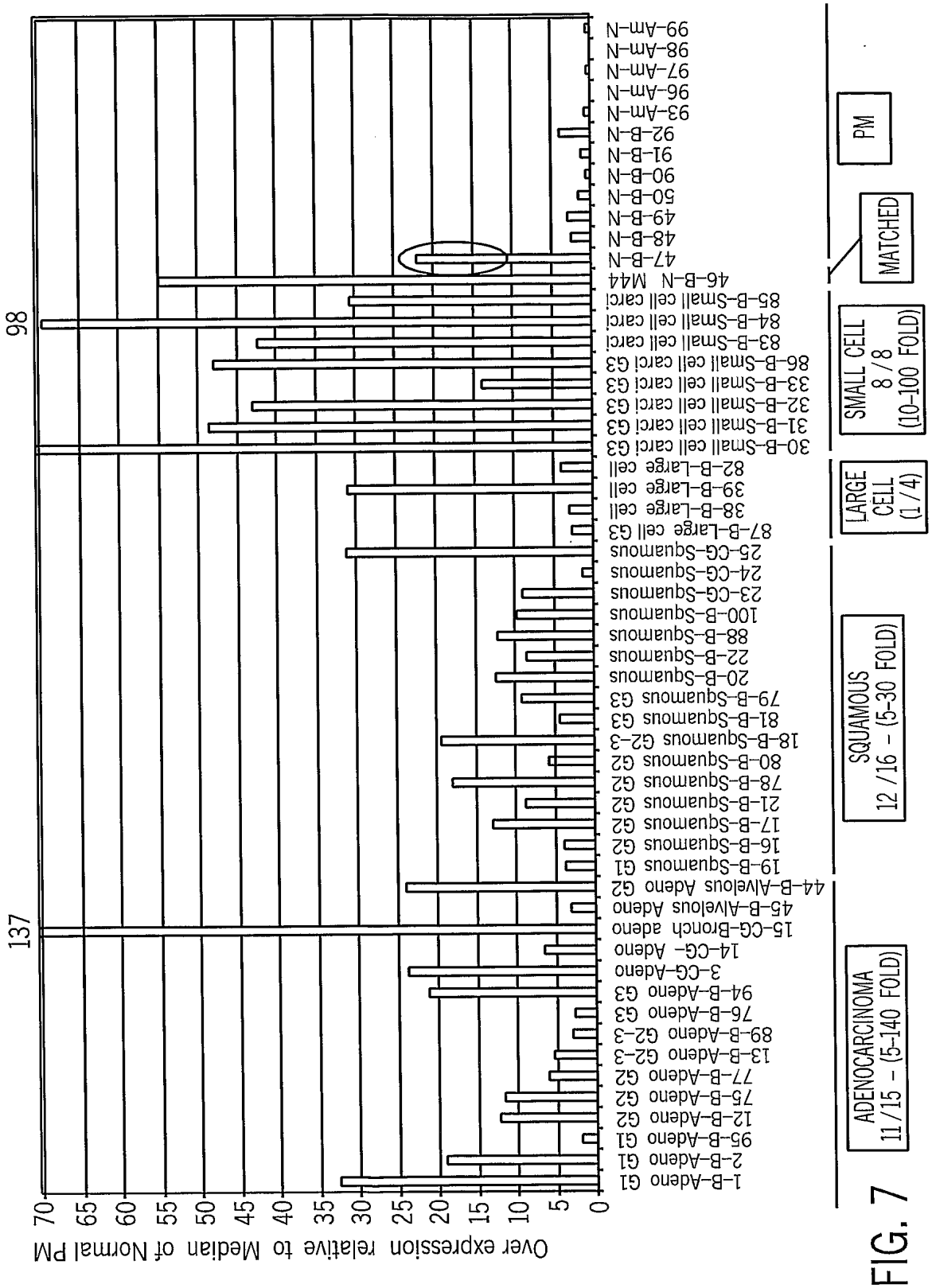

Over expression relative to Median of Normal PM
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
137
98
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N
ADENOCARCINOMA
11/15 – (5-140 FOLD)
SQUAMOUS
12/16 – (5-30 FOLD)
LARGE CELL (1/4)
SMALL CELL
8/8
(10-100 FOLD)
MATCHED
PM

FIG. 7

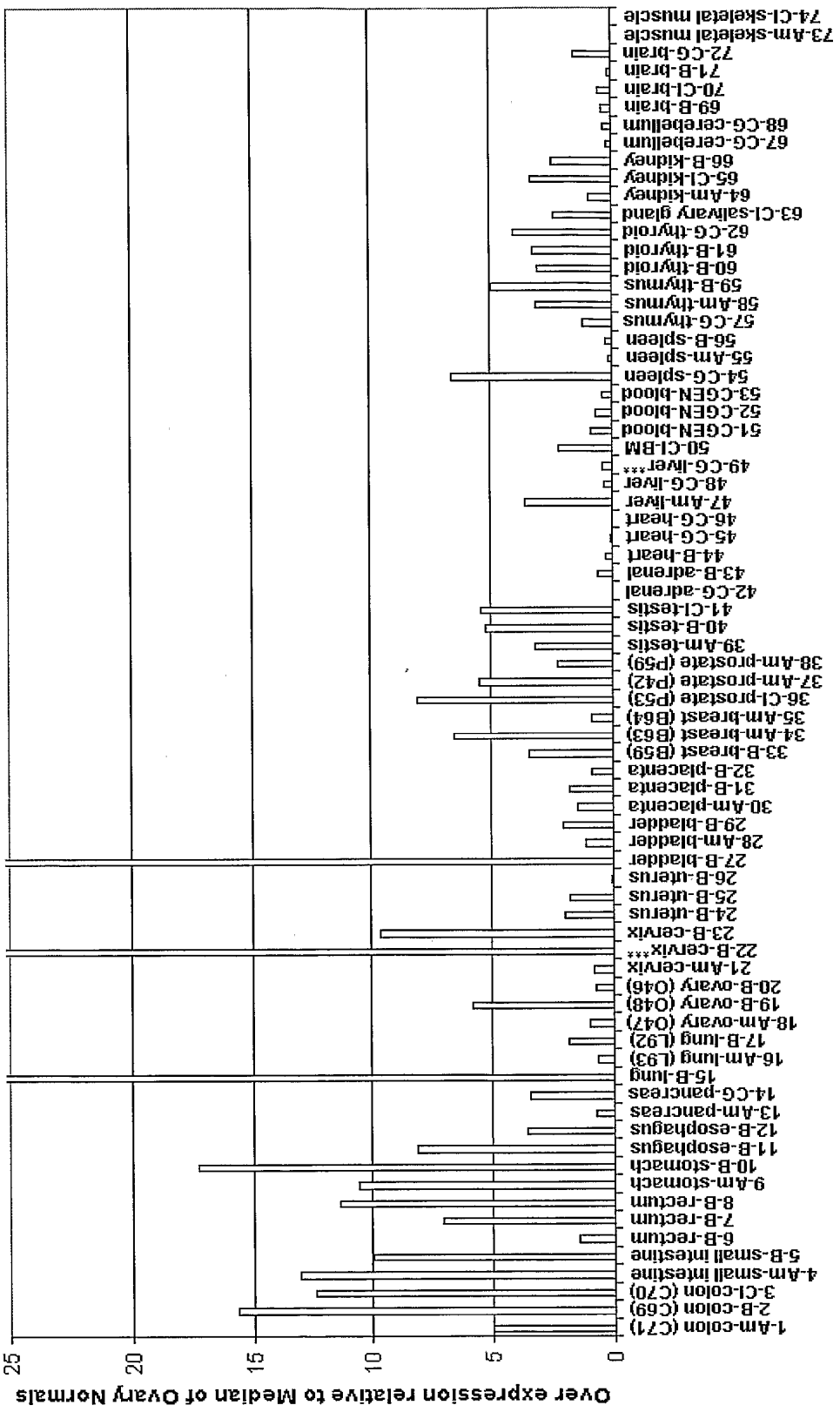

Over expression relative to Median of Ovary Normals
0
5
10
15
20
25
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 8

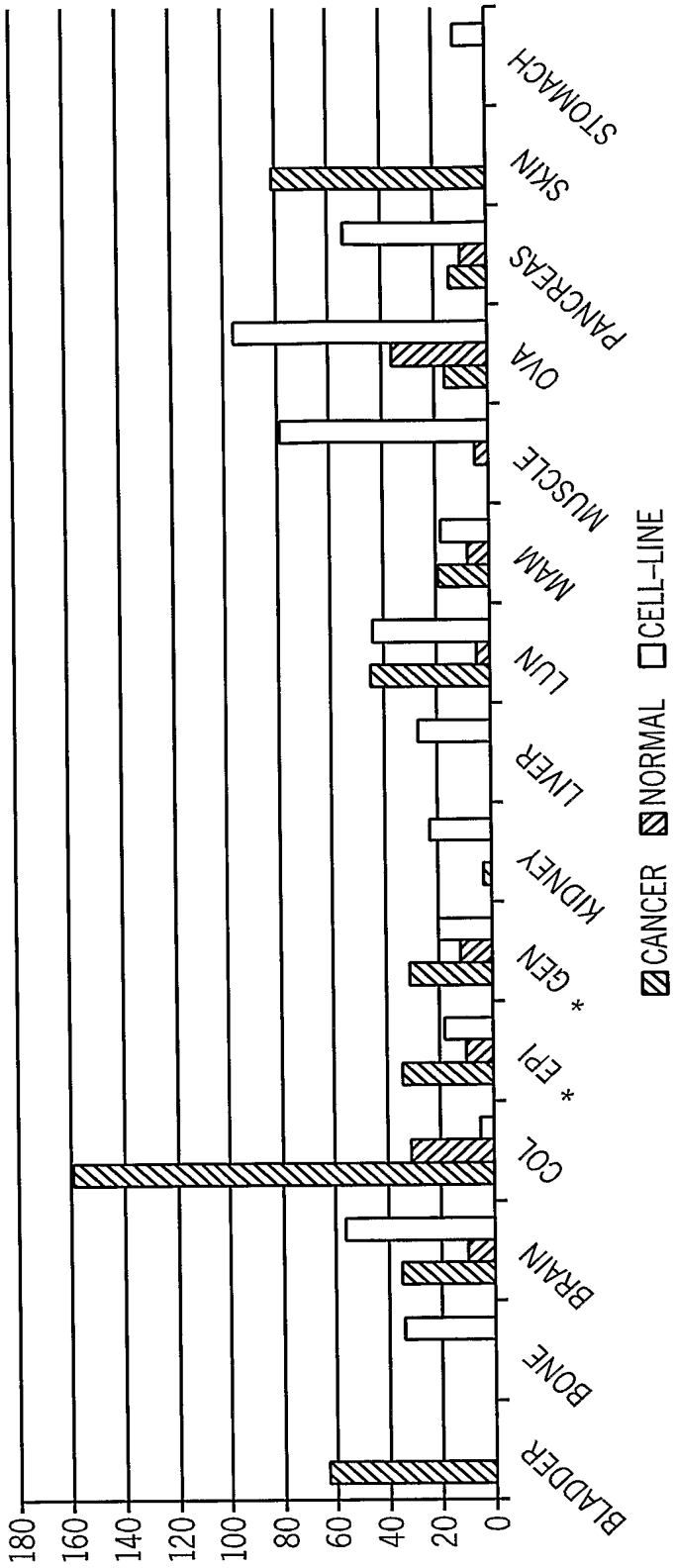

STOMACH
SKIN
PANCREAS
OVA
MUSCLE
MAM
LUN
LIVER
KIDNEY
* GEN
* EPI
COL
BRAIN
BONE
BLADDER
180
160
140
120
100
80
60
40
20
0
CANCER
NORMAL
CELL-LINE

FIG. 9

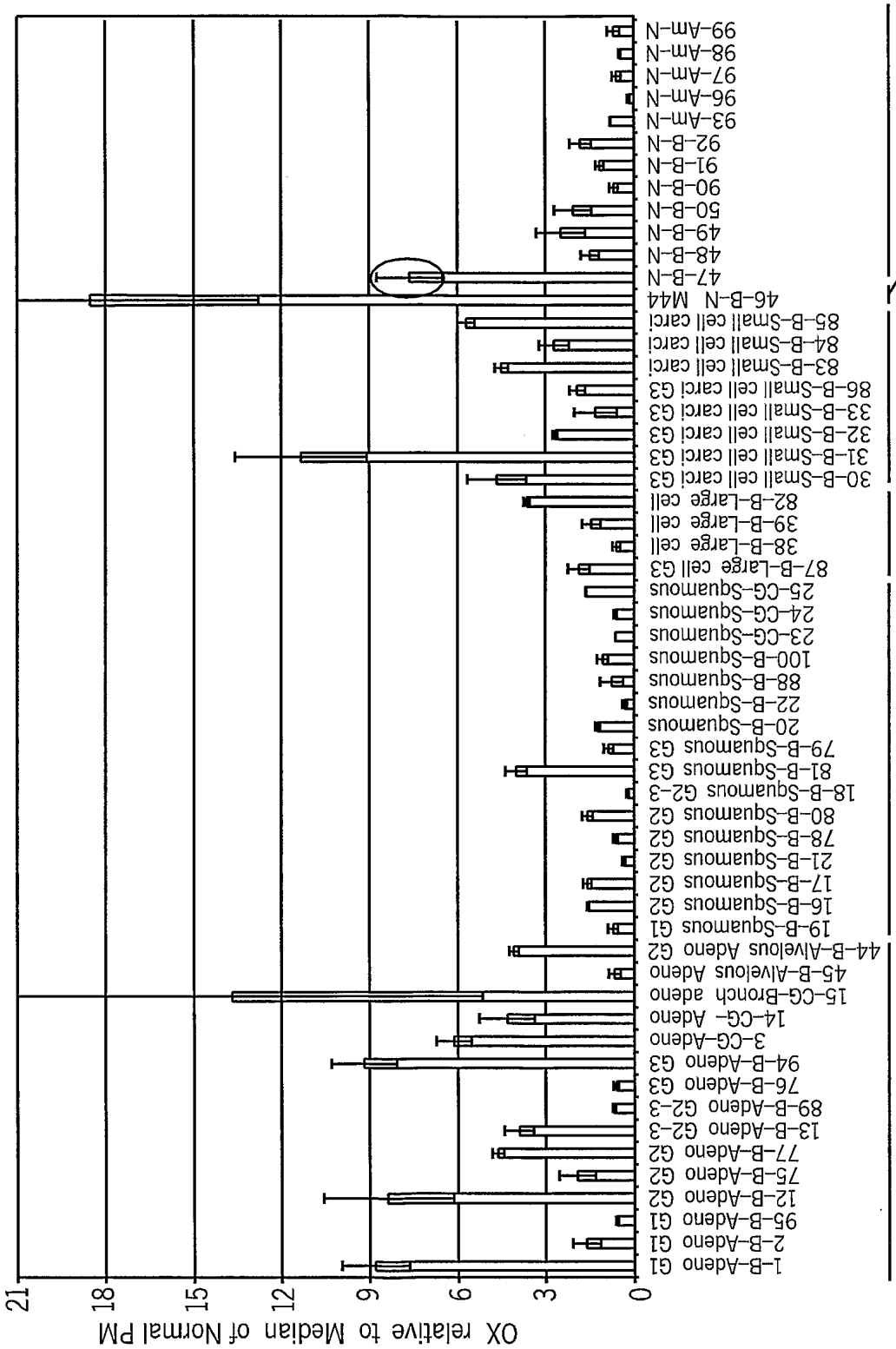

OX relative to Median of Normal PM
0
3
6
9
12
15
18
21
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N
ADENOCARCINOMA 9 /15 3-10 FOLD
SQUAMOUS
LARGE CELL
SMALL CELL 4 /8 3-12 FOLD
MATCHED
PM

FIG. 10

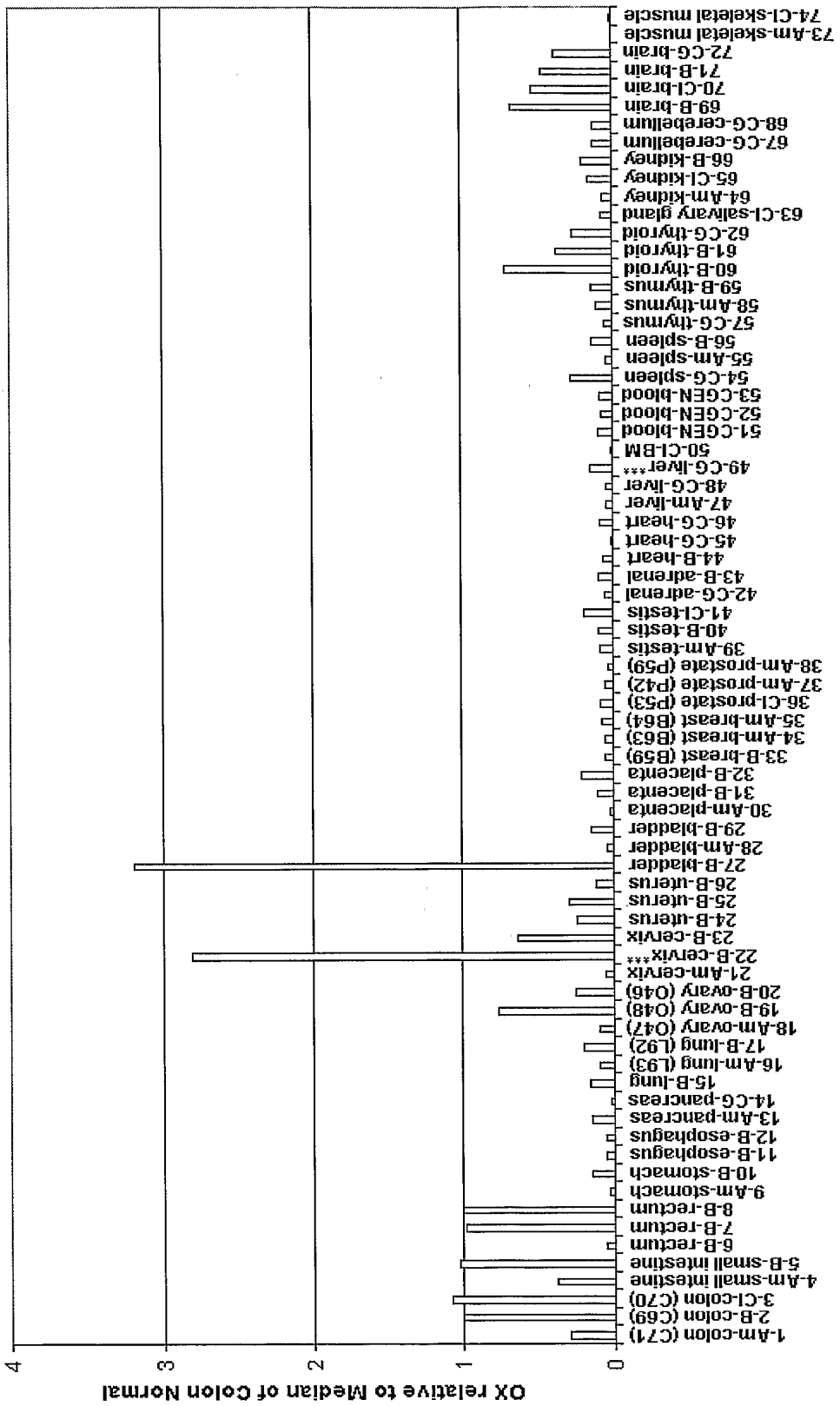

OX relative to Median of Colon Normal
0
1
2
3
4
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 11

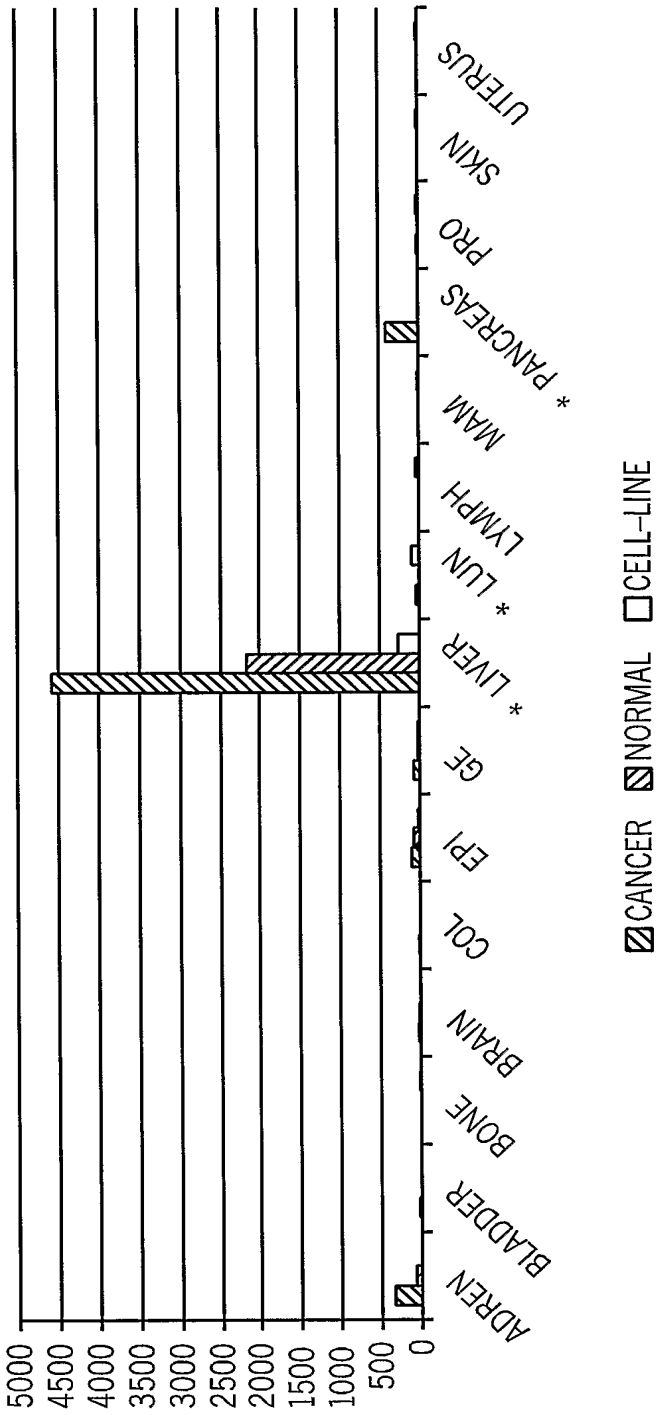

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
ADREN
BLADDER
BONE
BRAIN
COL
EPI
GE
* LIVER
* LUN
LYMPH
MAM
* PANCREAS
PRO
SKIN
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 12

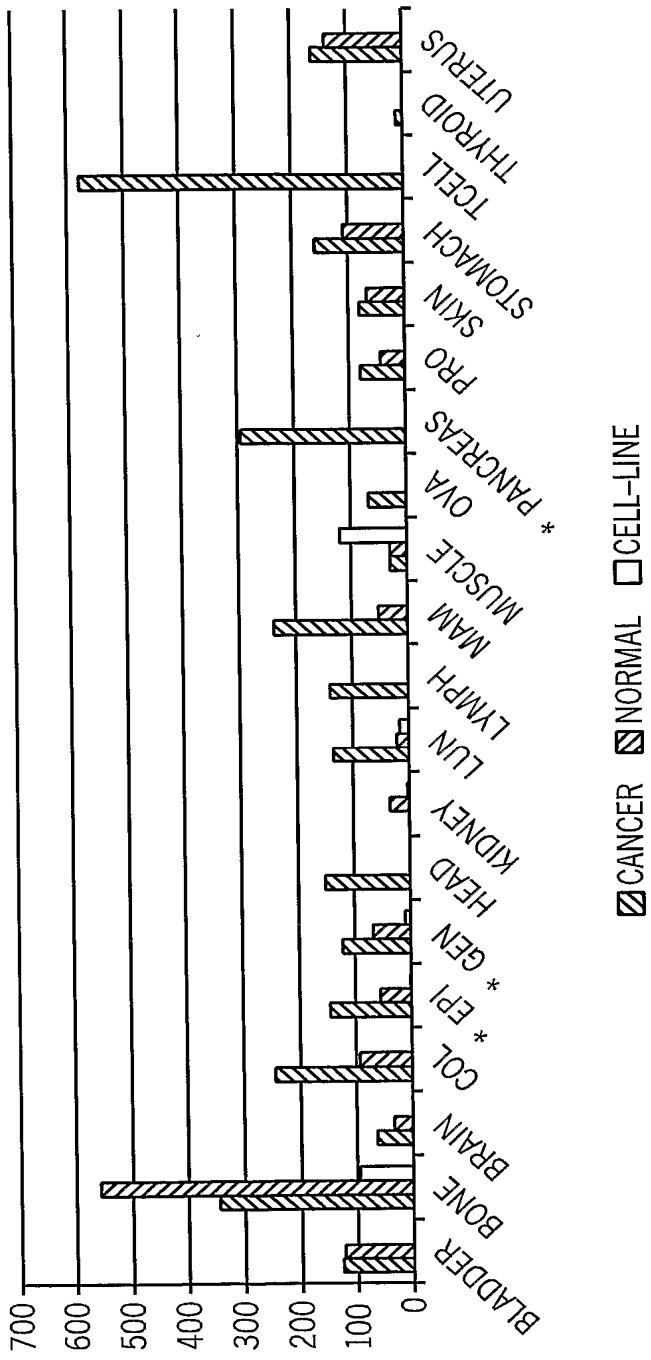

700
600
500
400
300
200
100
0
BLADDER
BONE
BRAIN
COL
* EPI
* GEN
HEAD
KIDNEY
LUN
LYMPH
MAM
MUSCLE
OVA
* PANCREAS
PRO
SKIN
STOMACH
TCELL
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 13

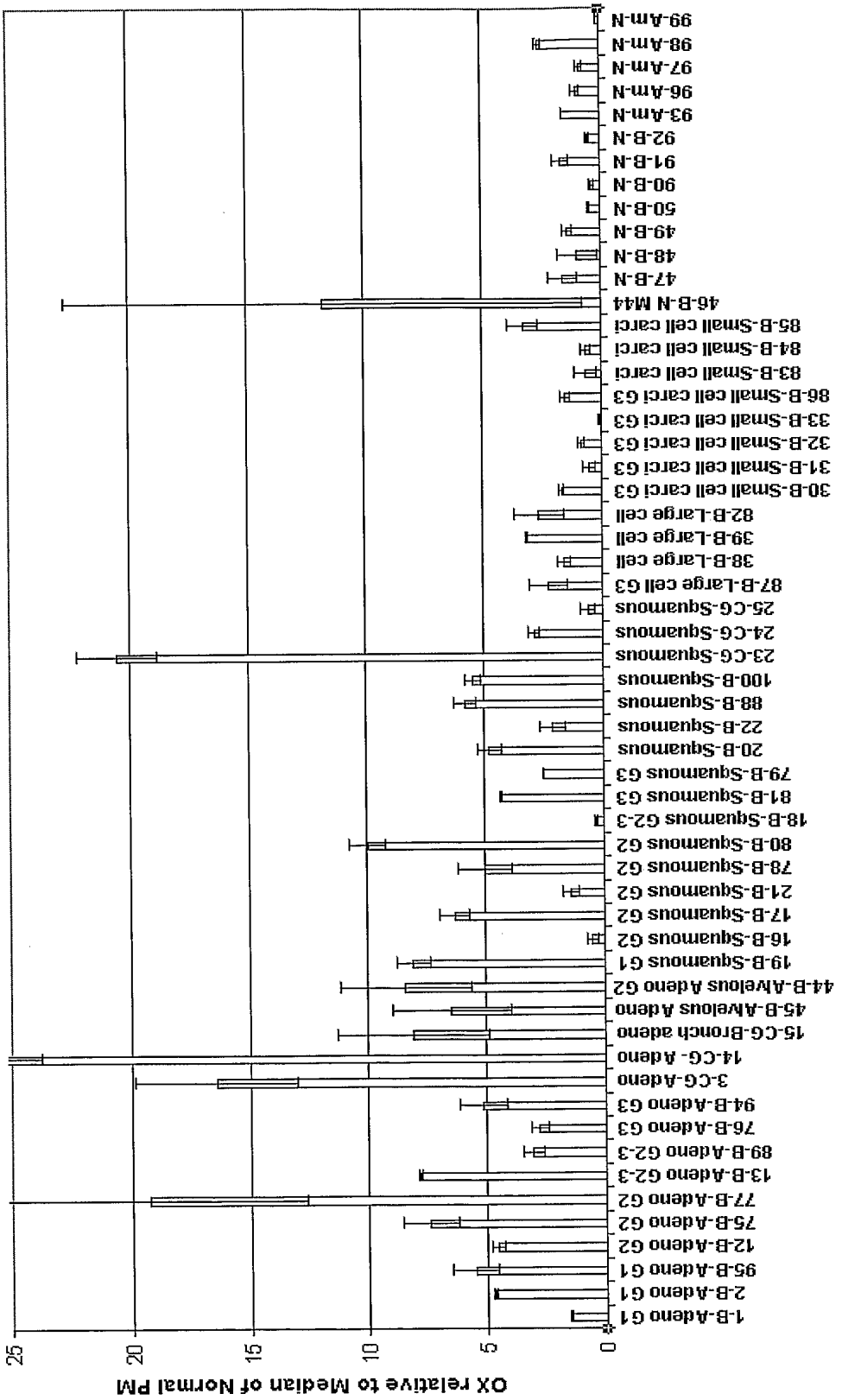

OX relative to Median of Normal PM
0
5
10
15
20
25
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 14

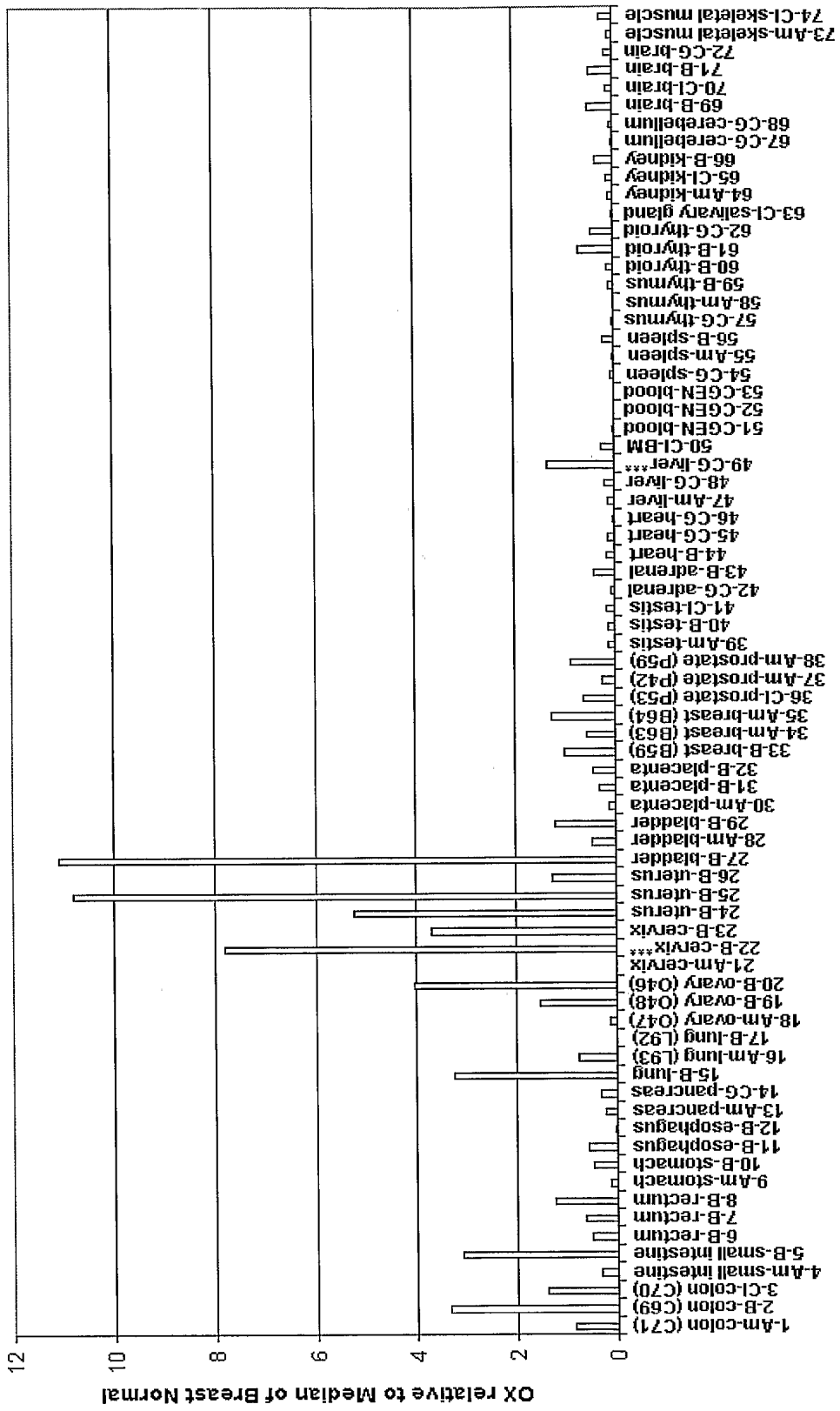

OX relative to Median of Breast Normal
0
2
4
6
8
10
12
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 15

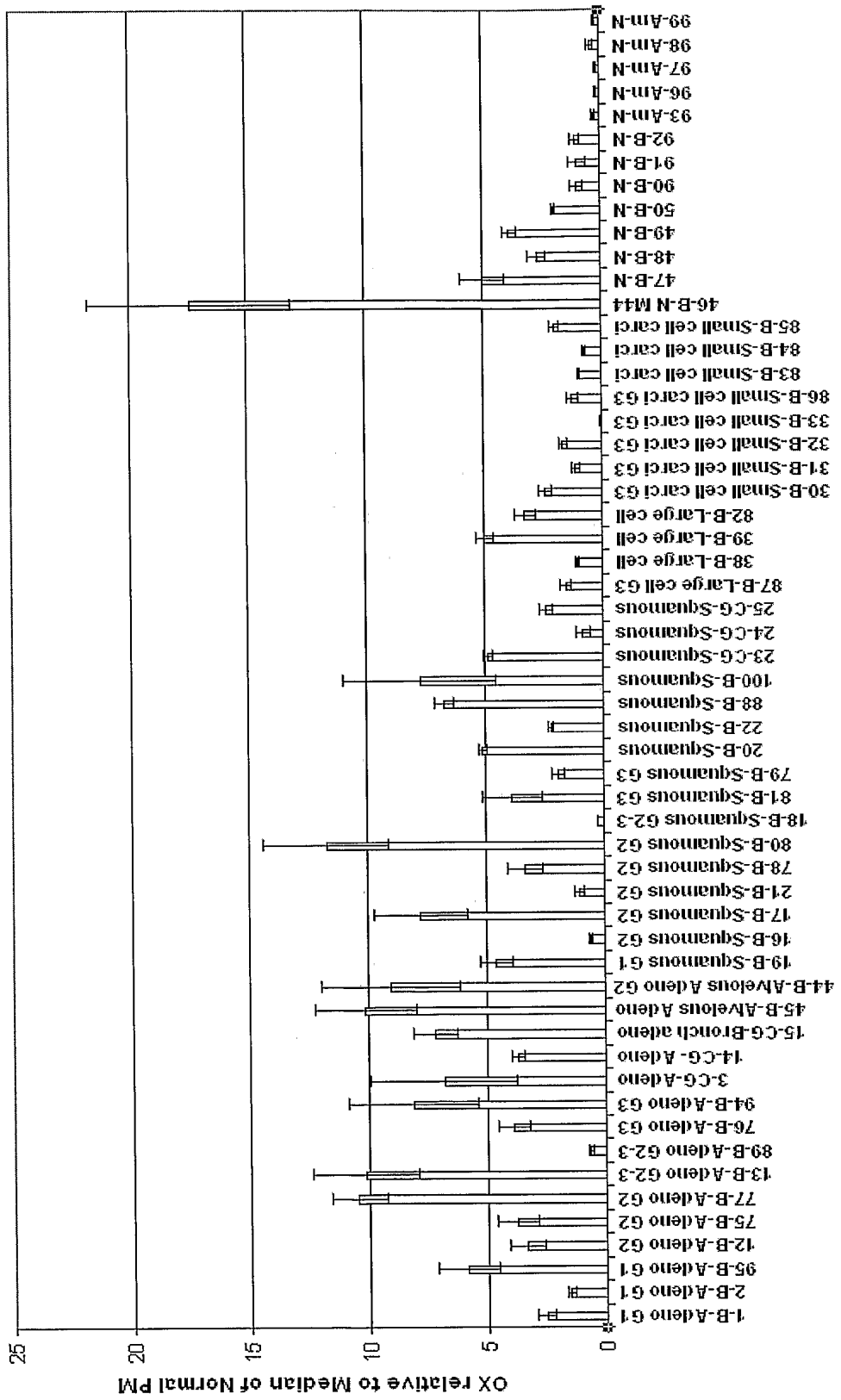

OX relative to Median of Normal PM
0
5
10
15
20
25
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 16

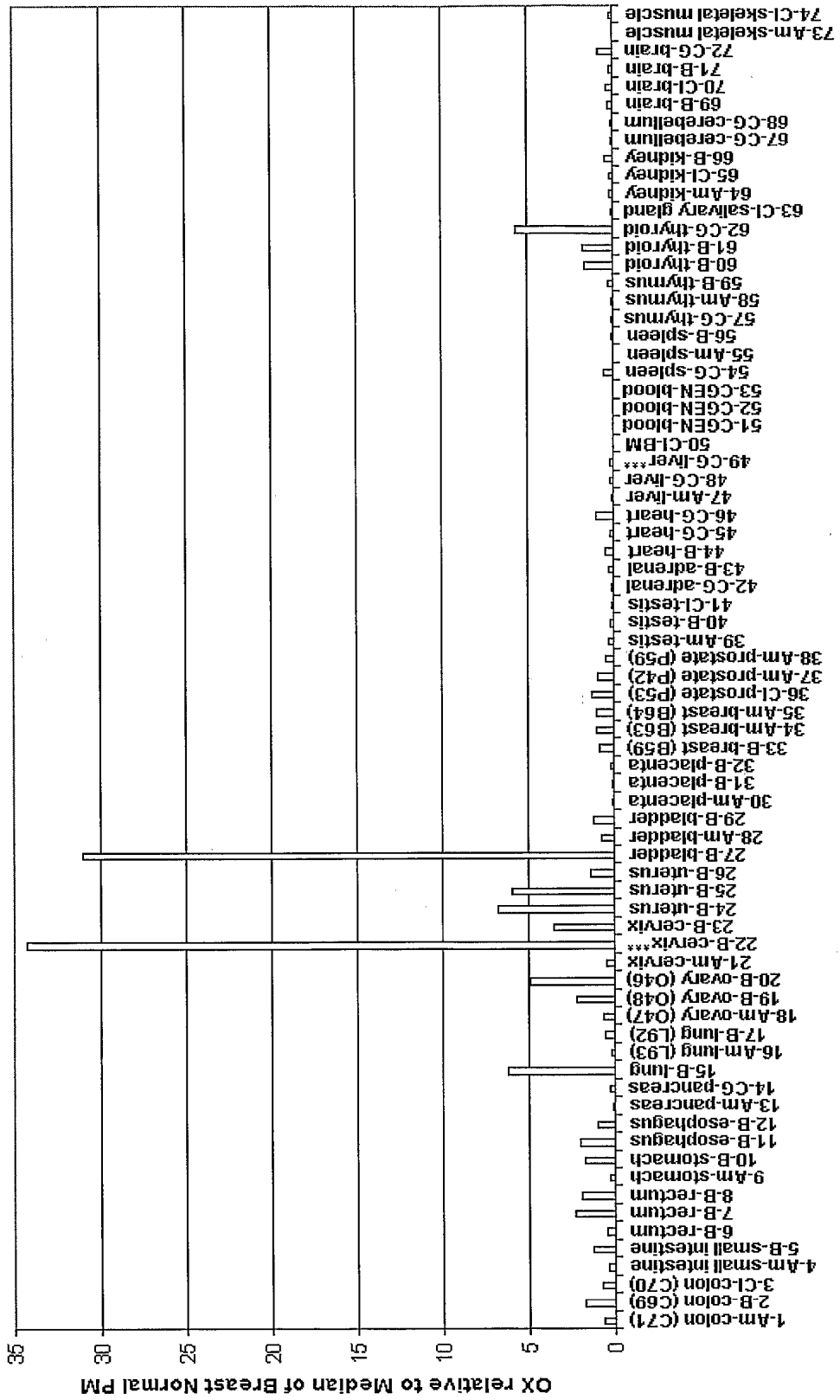

OX relative to Median of Breast Normal PM
0
5
10
15
20
25
30
35
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 17

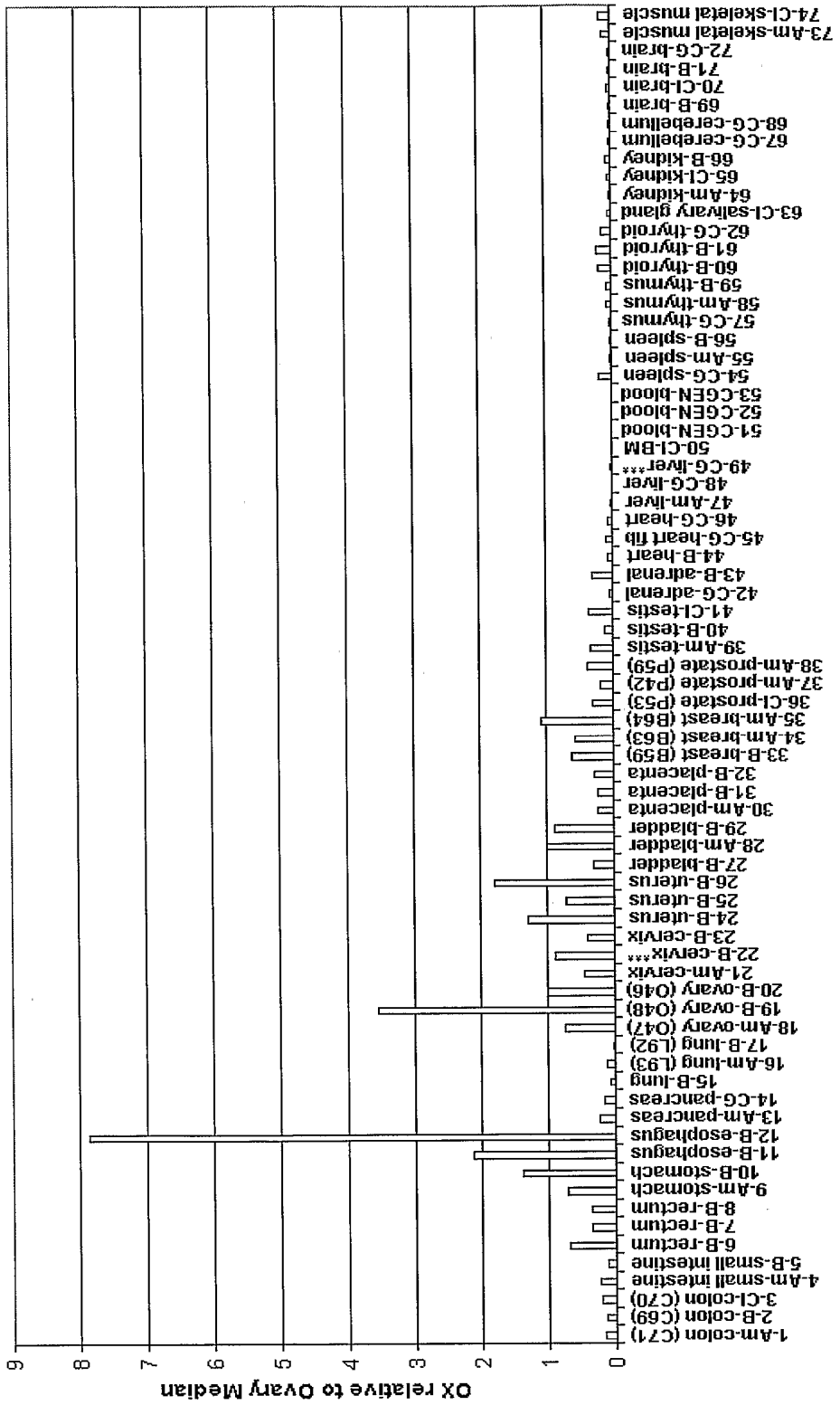

OX relative to Ovary Median
0
1
2
3
4
5
6
7
8
9
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart f/b
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 18

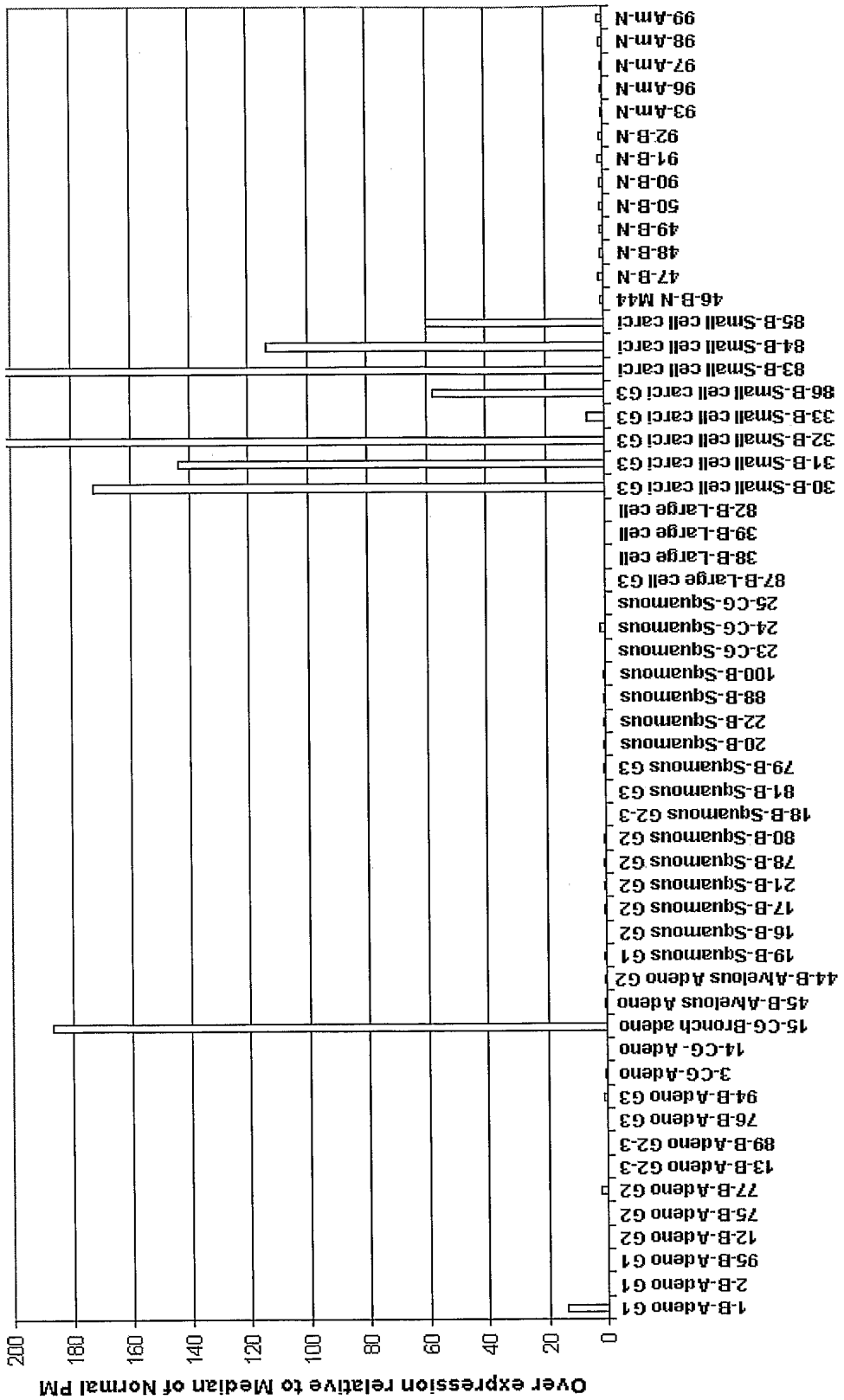

Over expression relative to Median of Normal PM
0
20
40
60
80
100
120
140
160
180
200
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 19

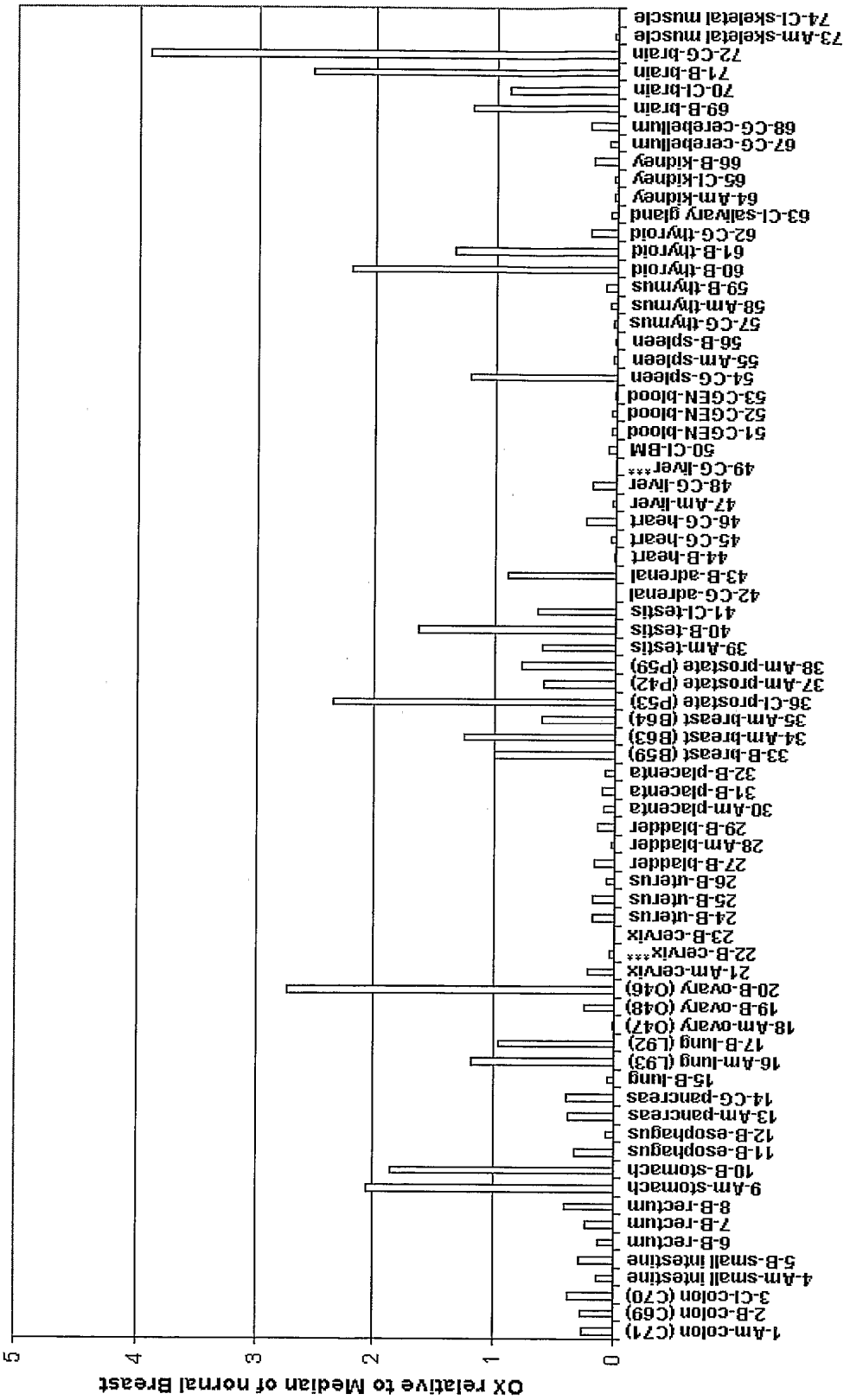

OX relative to Median of normal Breast
0
1
2
3
4
5
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 20

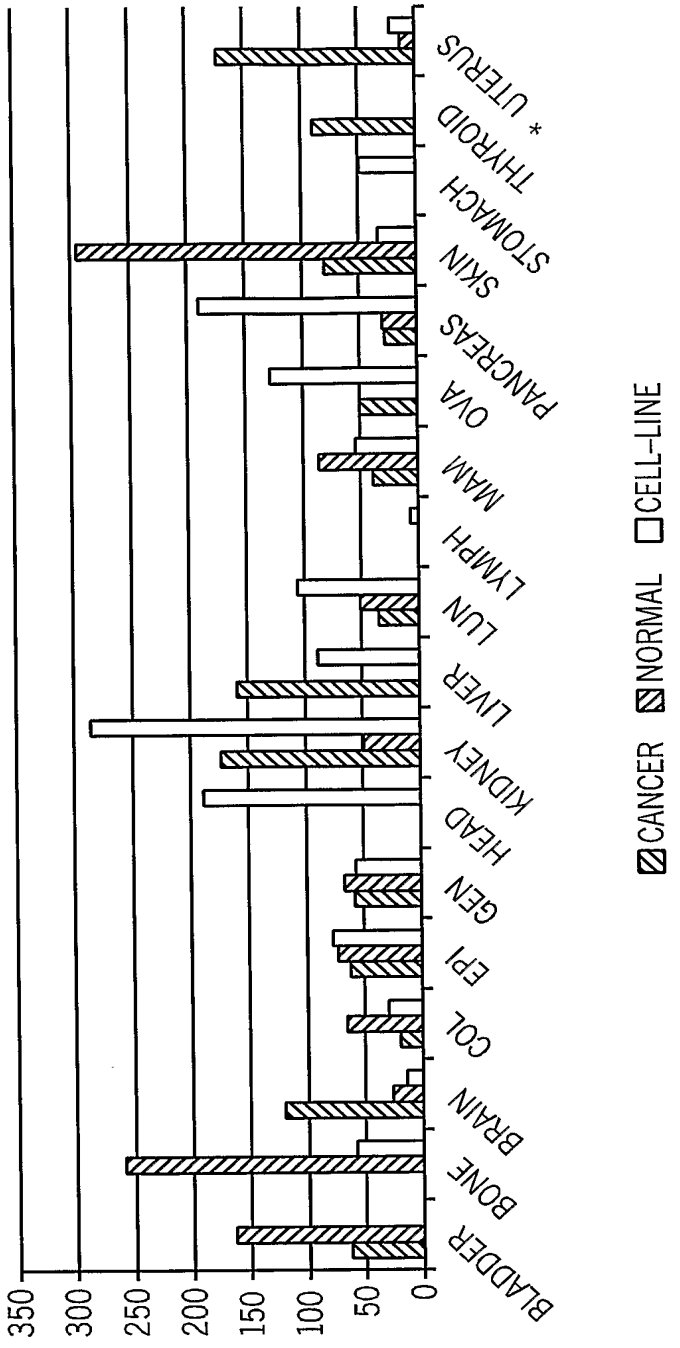

350
300
250
200
150
100
50
0
BLADDER
BONE
BRAIN
COL
EPI
GEN
HEAD
KIDNEY
LIVER
LUN
LYMPH
MAM
OVA
PANCREAS
SKIN
STOMACH
THYROID
* UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 21

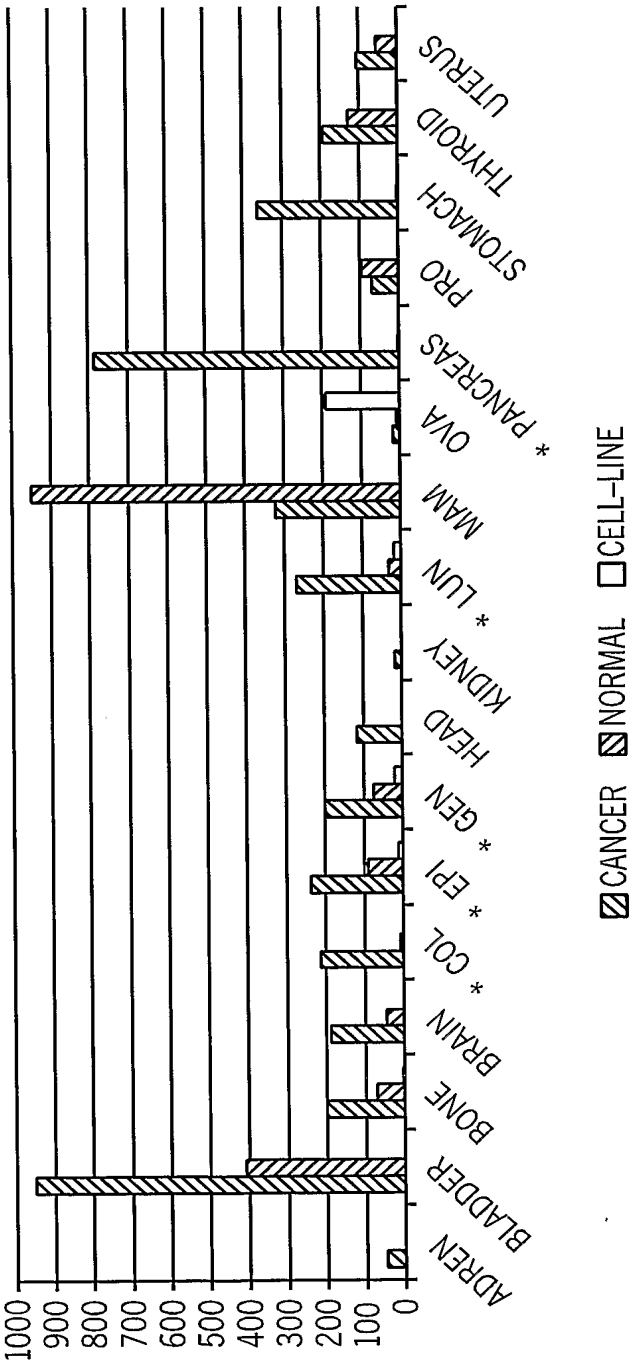

1000
900
800
700
600
500
400
300
200
100
0
ADREN
BLADDER
BONE
BRAIN
* COL
* EPI
* GEN
HEAD
KIDNEY
* LUN
MAM
OVA
* PANCREAS
PRO
STOMACH
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 22

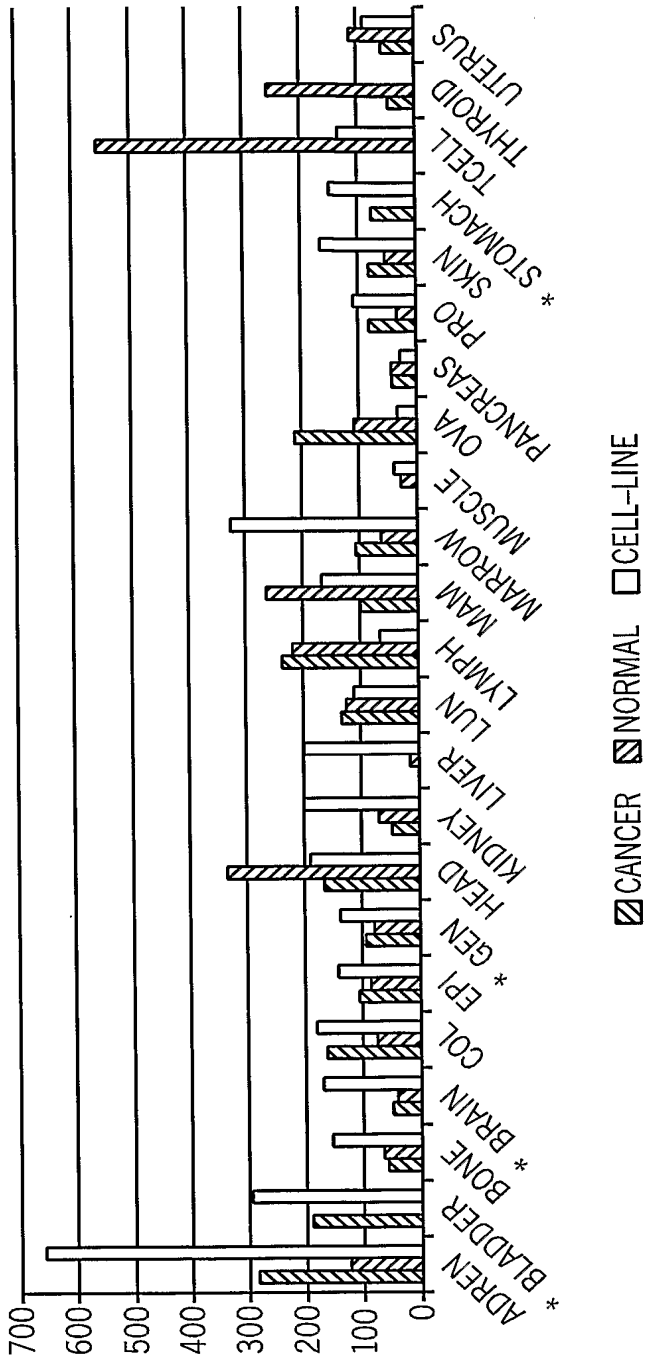

700
600
500
400
300
200
100
0
* ADREN
* BLADDER
BONE
* BRAIN
COL
EPI
* GEN
HEAD
KIDNEY
LIVER
LUN
LYMPH
MAM
MARROW
MUSCLE
OVA
PANCREAS
PRO
SKIN
* STOMACH
TCELL
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 23

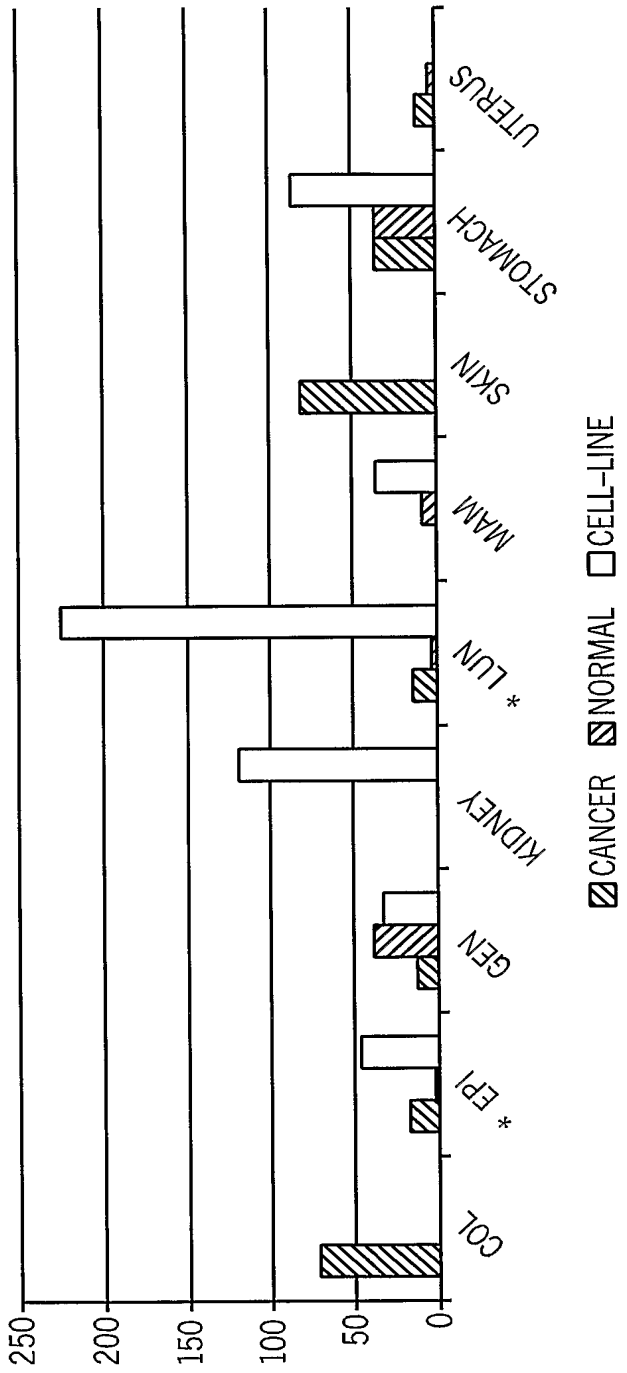

250
200
150
100
50
0
COL
* EPI
GEN
KIDNEY
* LUN
MAM
SKIN
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 24

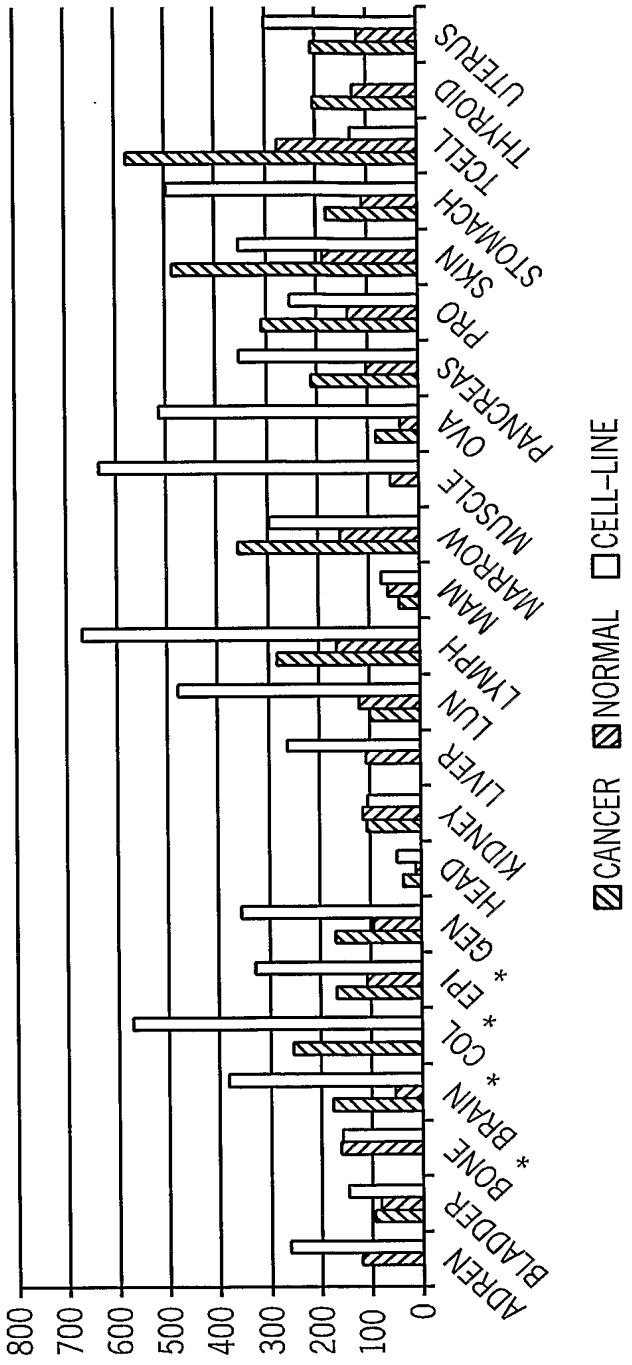

800
700
600
500
400
300
200
100
0
ADREN
BLADDER
BONE
* BRAIN
* COL
* EPI
* GEN
HEAD
KIDNEY
LIVER
LUN
LYMPH
MAM
MARROW
MUSCLE
OVA
PANCREAS
PRO
SKIN
STOMACH
TCELL
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 25

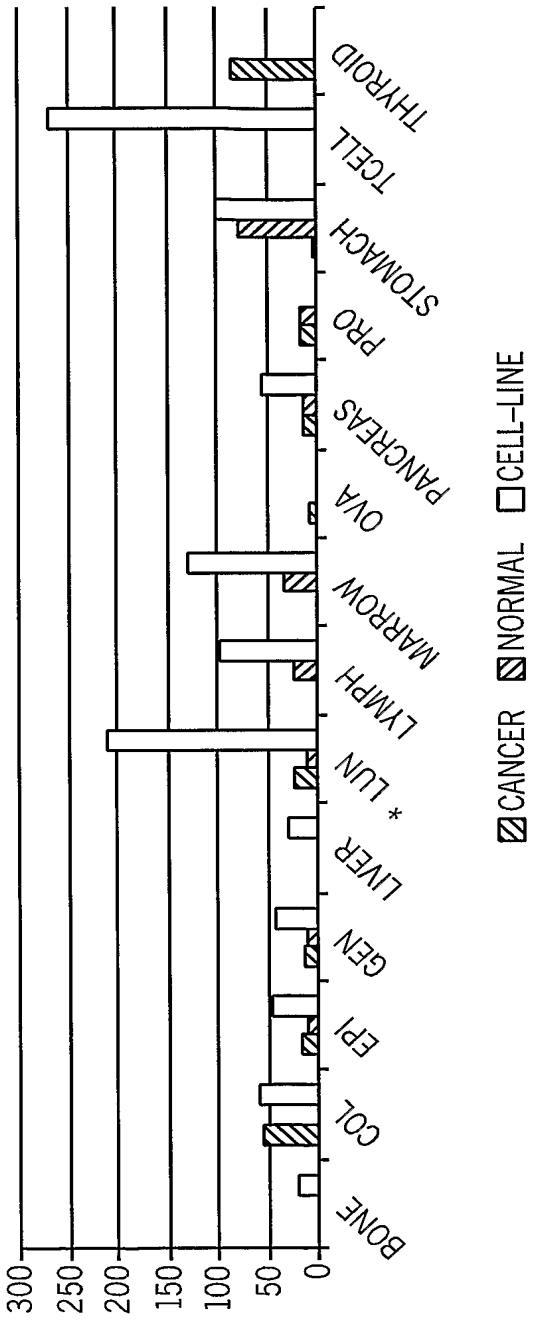

300
250
200
150
100
50
0
BONE
COL
EPI
GEN
LIVER
* LUN
LYMPH
MARROW
OVA
PANCREAS
PRO
STOMACH
TCELL
THYROID
CANCER
NORMAL
CELL-LINE

FIG. 26

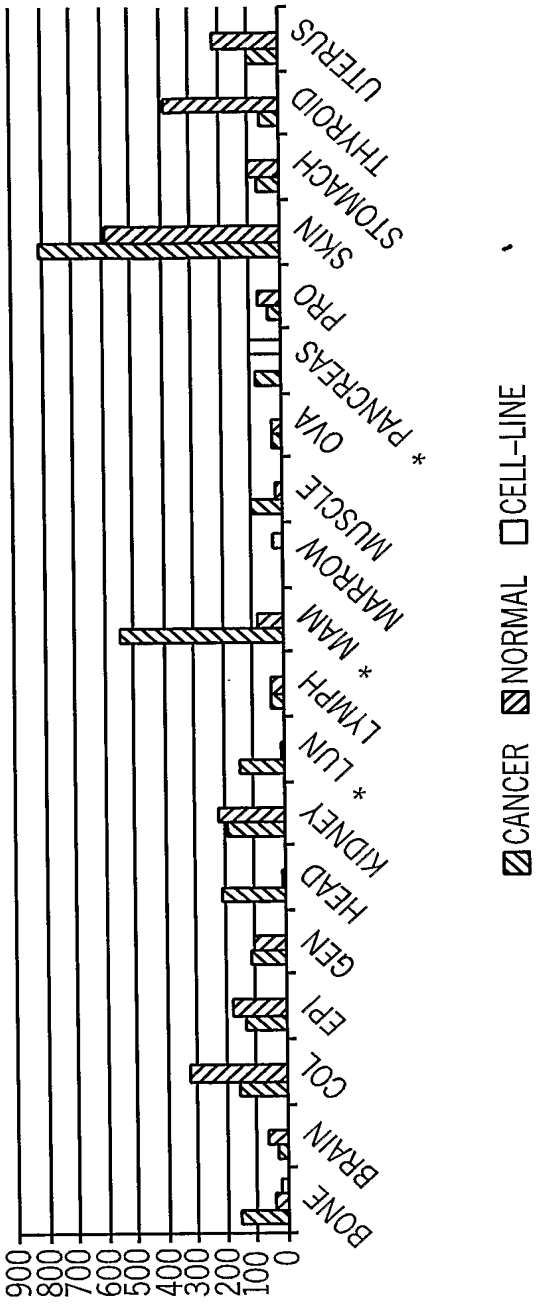

900
800
700
600
500
400
300
200
100
0
BONE
BRAIN
COL
EPI
GEN
HEAD
KIDNEY
* LUN
LYMPH
* MAM
MARROW
MUSCLE
OVA
* PANCREAS
PRO
SKIN
STOMACH
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 27

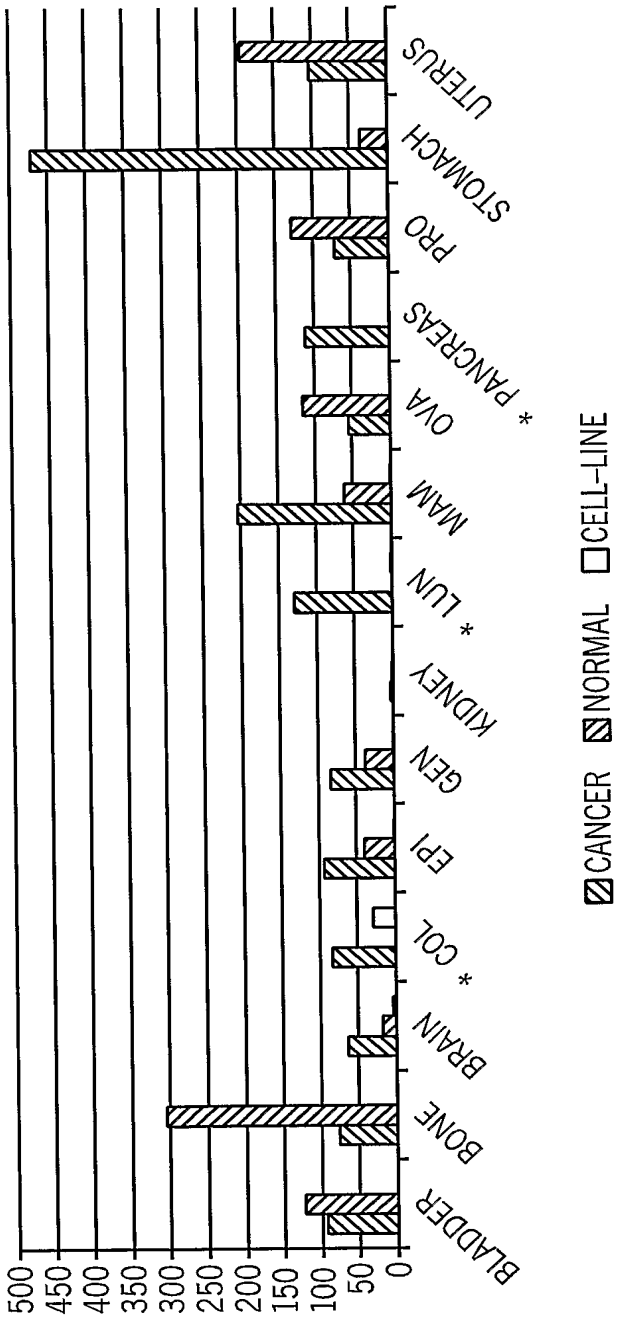

500
450
400
350
300
250
200
150
100
50
0
BLADDER
BONE
BRAIN
* COL
EPI
GEN
KIDNEY
* LUN
MAM
OVA
* PANCREAS
PRO
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 28

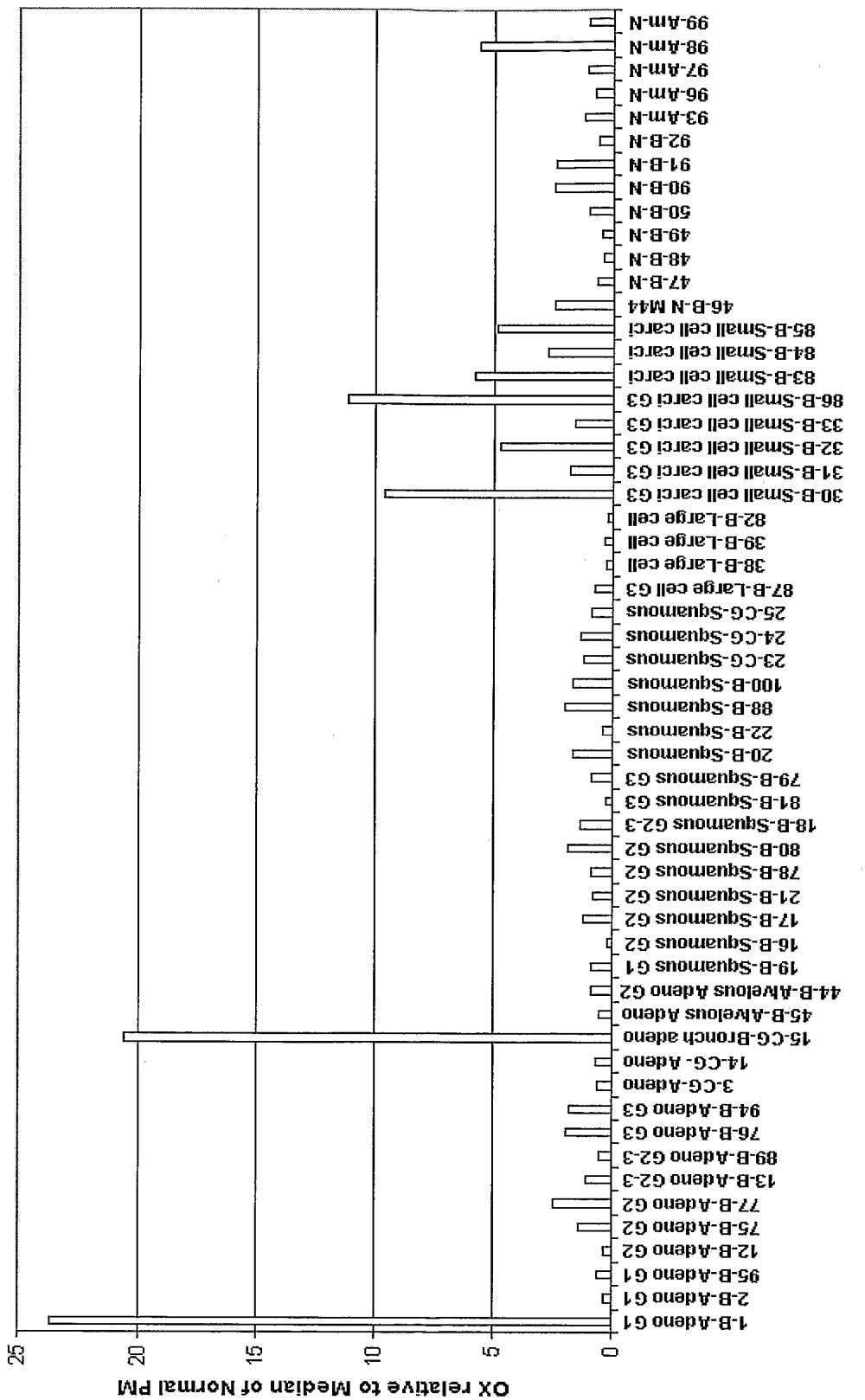

OX relative to Median of Normal PM
0
5
10
15
20
25
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 29

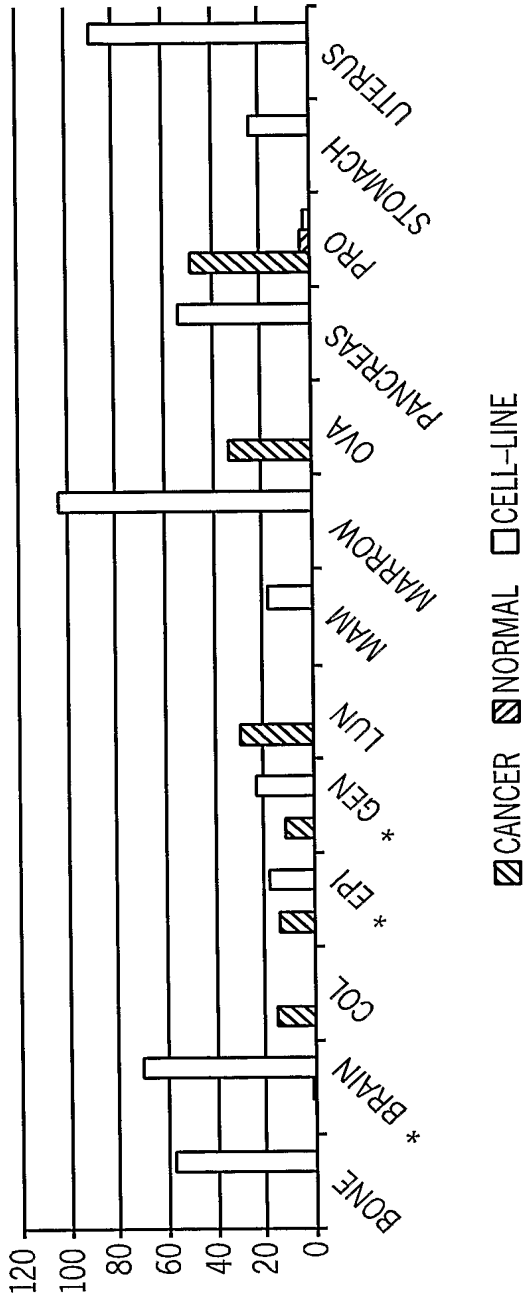

120
100
80
60
40
20
0
BONE
* BRAIN
COL
* EPI
* GEN
LUN
MAM
MARROW
OVA
PANCREAS
PRO
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 30

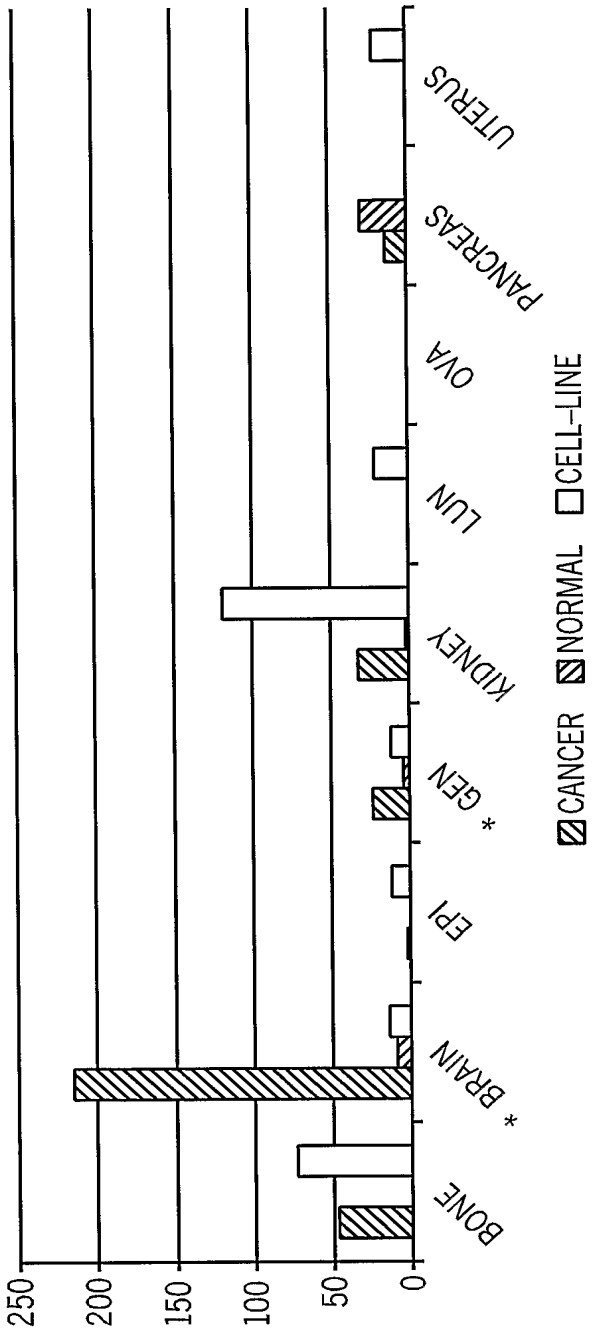

250
200
150
100
50
0
BONE
* BRAIN
EPI
* GEN
KIDNEY
LUN
OVA
PANCREAS
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 31

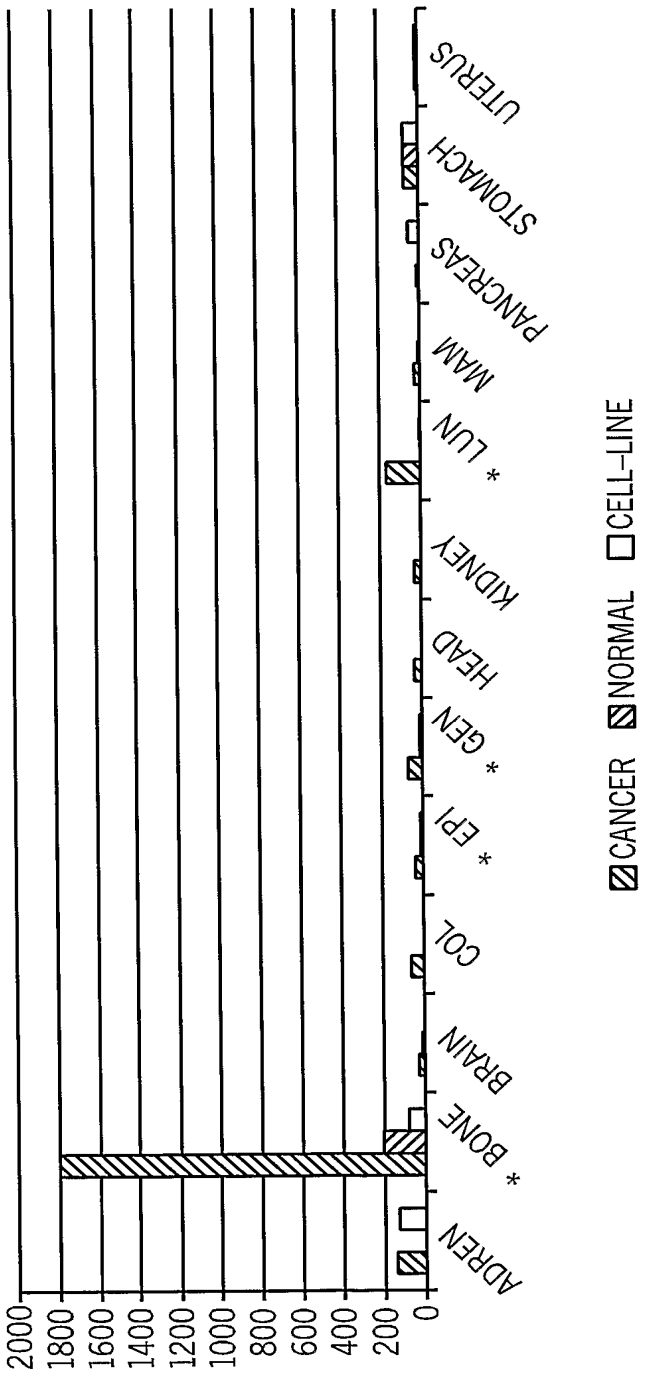

2000
1800
1600
1400
1200
1000
800
600
400
200
0
ADREN
* BONE
BRAIN
COL
* EPI
* GEN
HEAD
KIDNEY
* LUN
MAM
PANCREAS
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 32

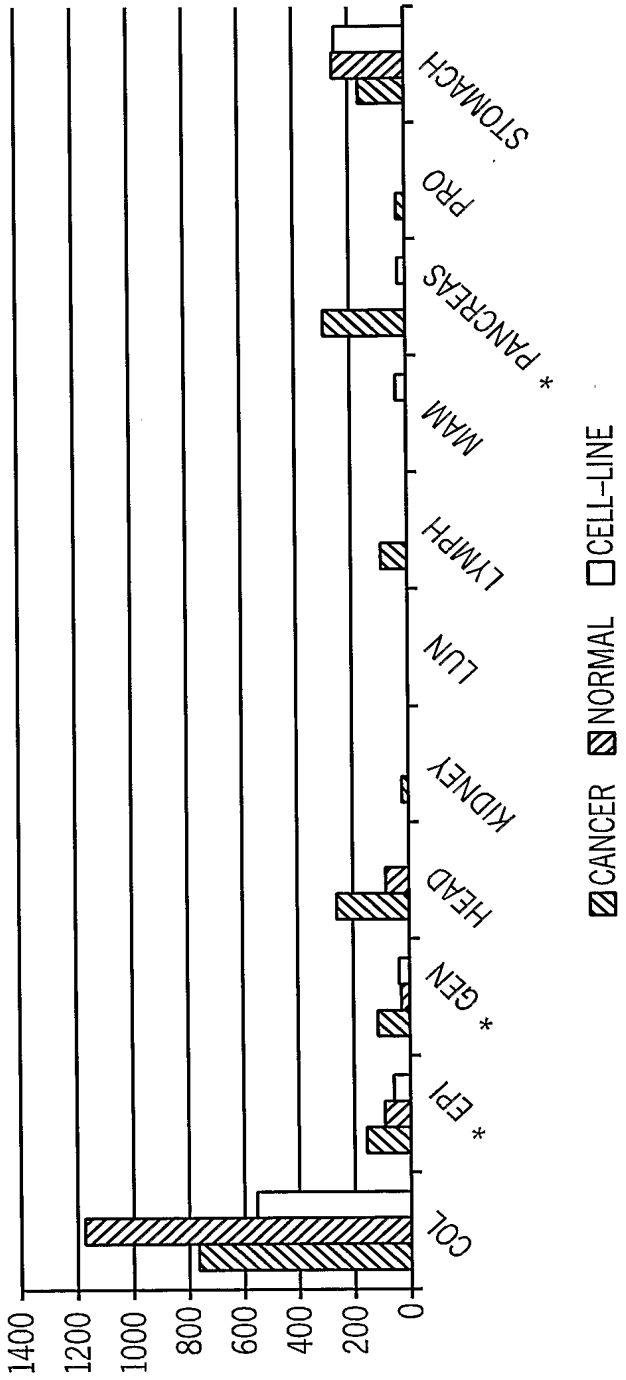

1400
1200
1000
800
600
400
200
0
COL
* EPI
* GEN
HEAD
KIDNEY
LUN
LYMPH
MAM
* PANCREAS
PRO
STOMACH
CANCER
NORMAL
CELL-LINE

FIG. 33

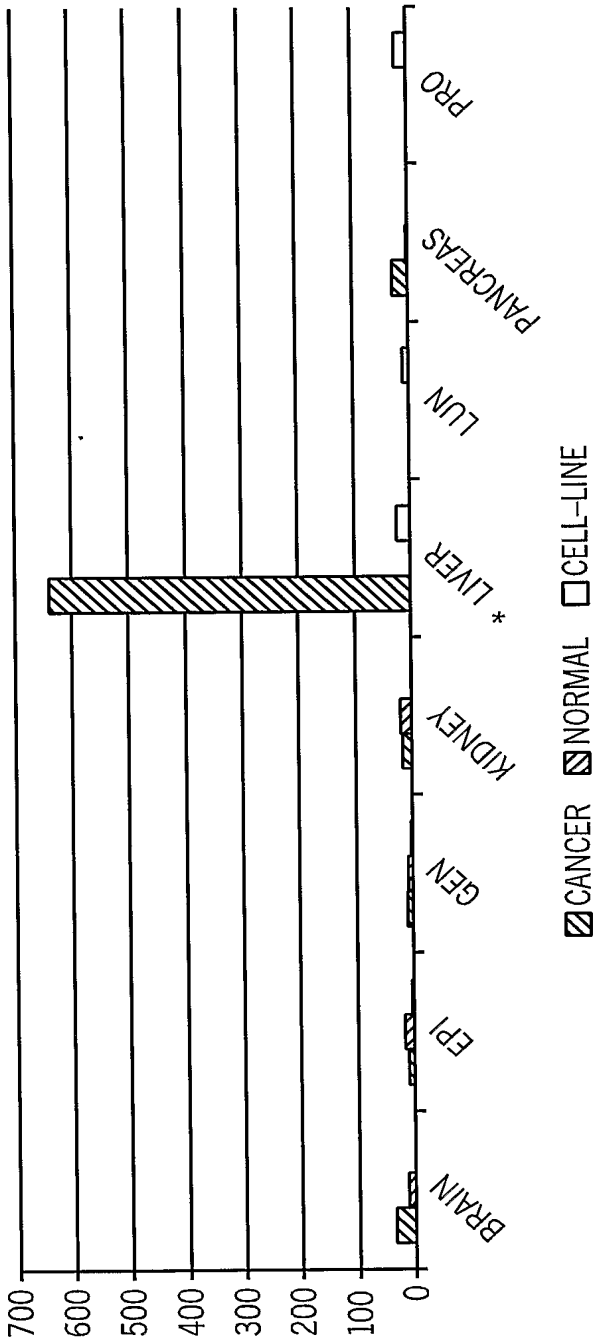

700
600
500
400
300
200
100
0
BRAIN
EPI
GEN
KIDNEY
* LIVER
LUN
PANCREAS
PRO
CANCER
NORMAL
CELL-LINE

FIG. 34

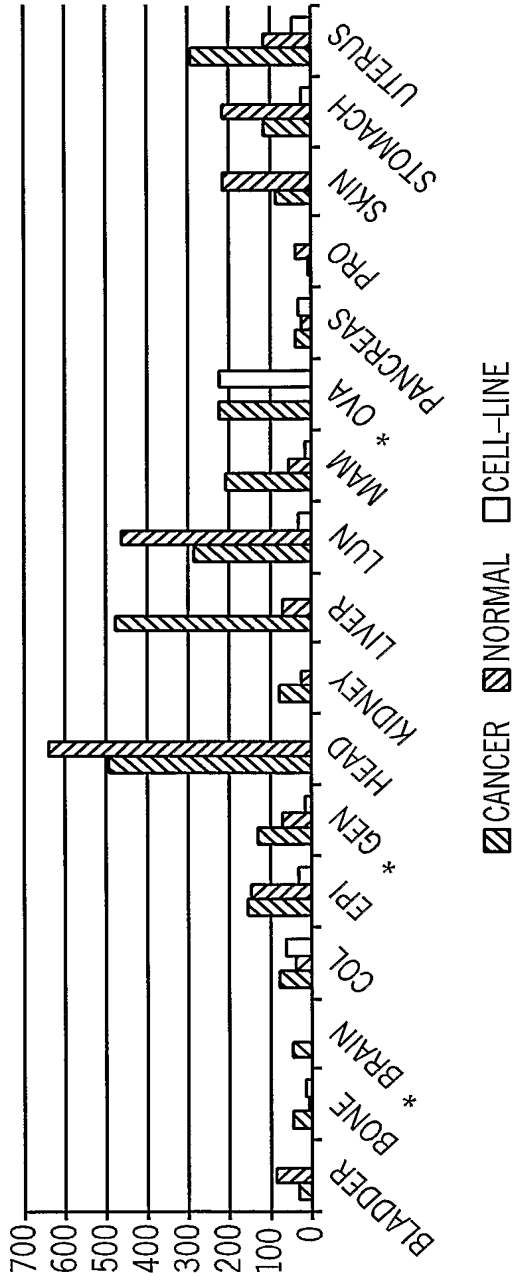

700
600
500
400
300
200
100
0
BLADDER
BONE
* BRAIN
COL
EPI
* GEN
HEAD
KIDNEY
LIVER
LUN
MAM
* OVA
PANCREAS
PRO
SKIN
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 35

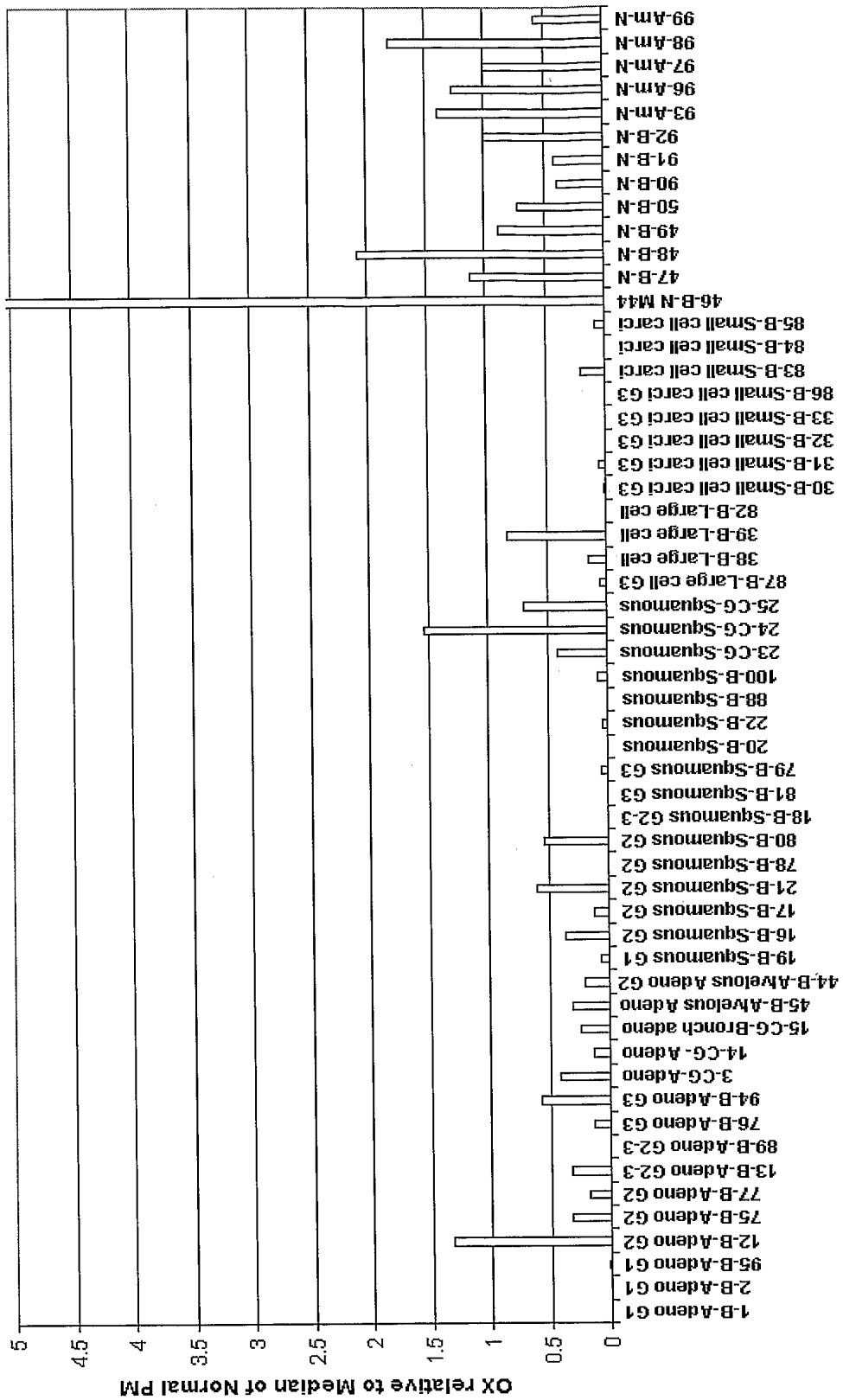

OX relative to Median of Normal PM
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 36

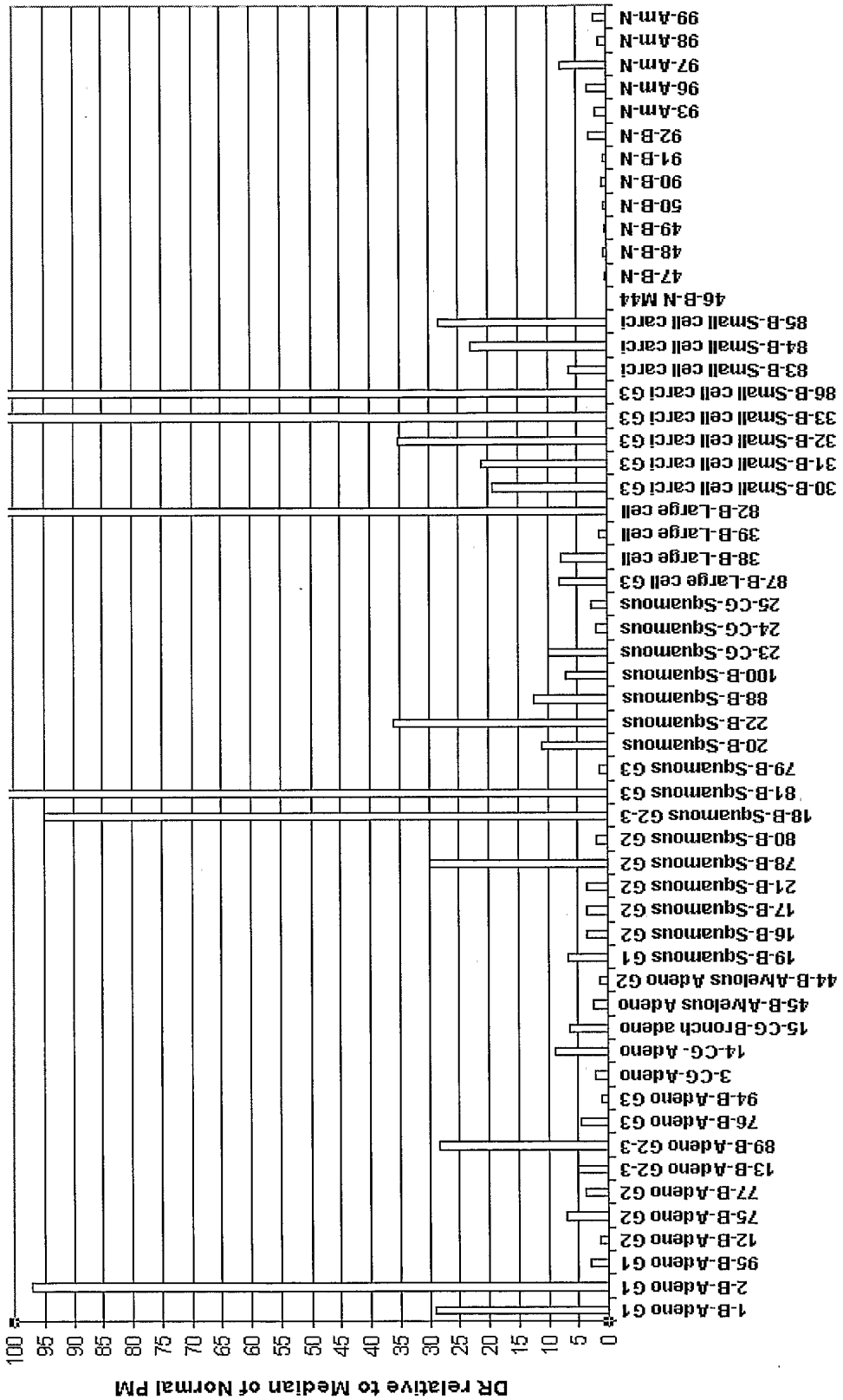

DR relative to Median of Normal PM
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 37

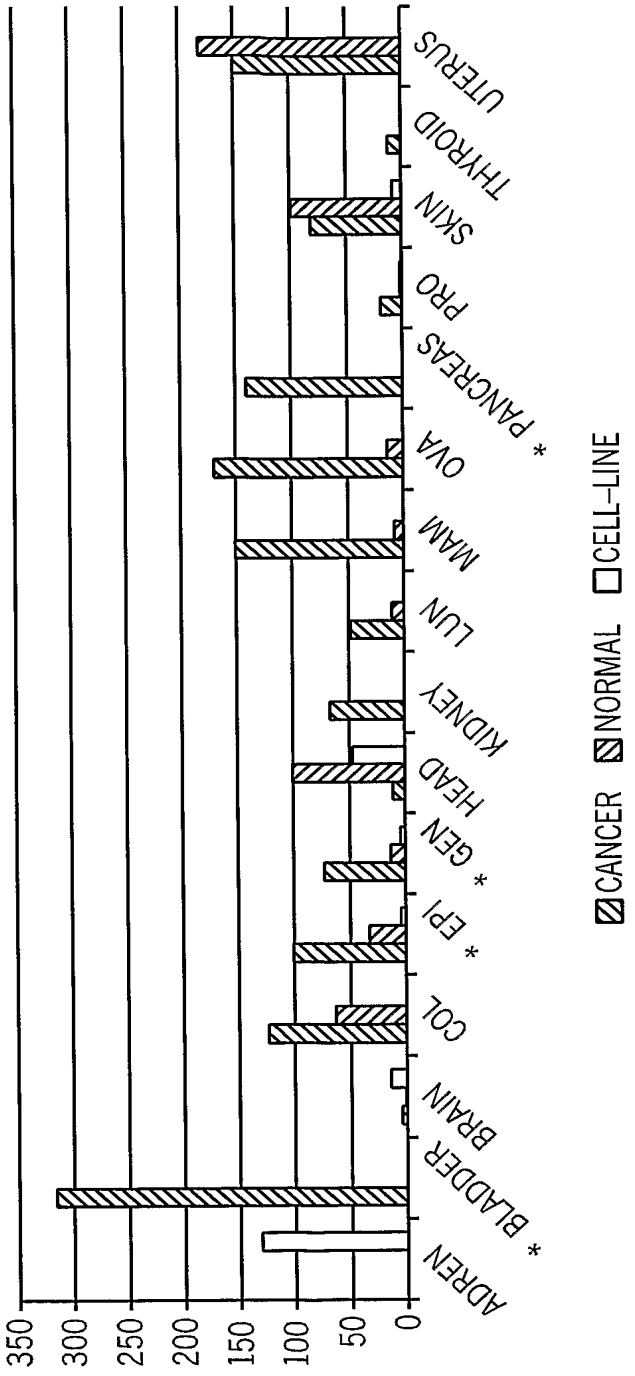

350
300
250
200
150
100
50
0
ADREN
* BLADDER
BRAIN
COL
* EPI
* GEN
HEAD
KIDNEY
LUN
MAM
OVA
* PANCREAS
PRO
SKIN
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 38

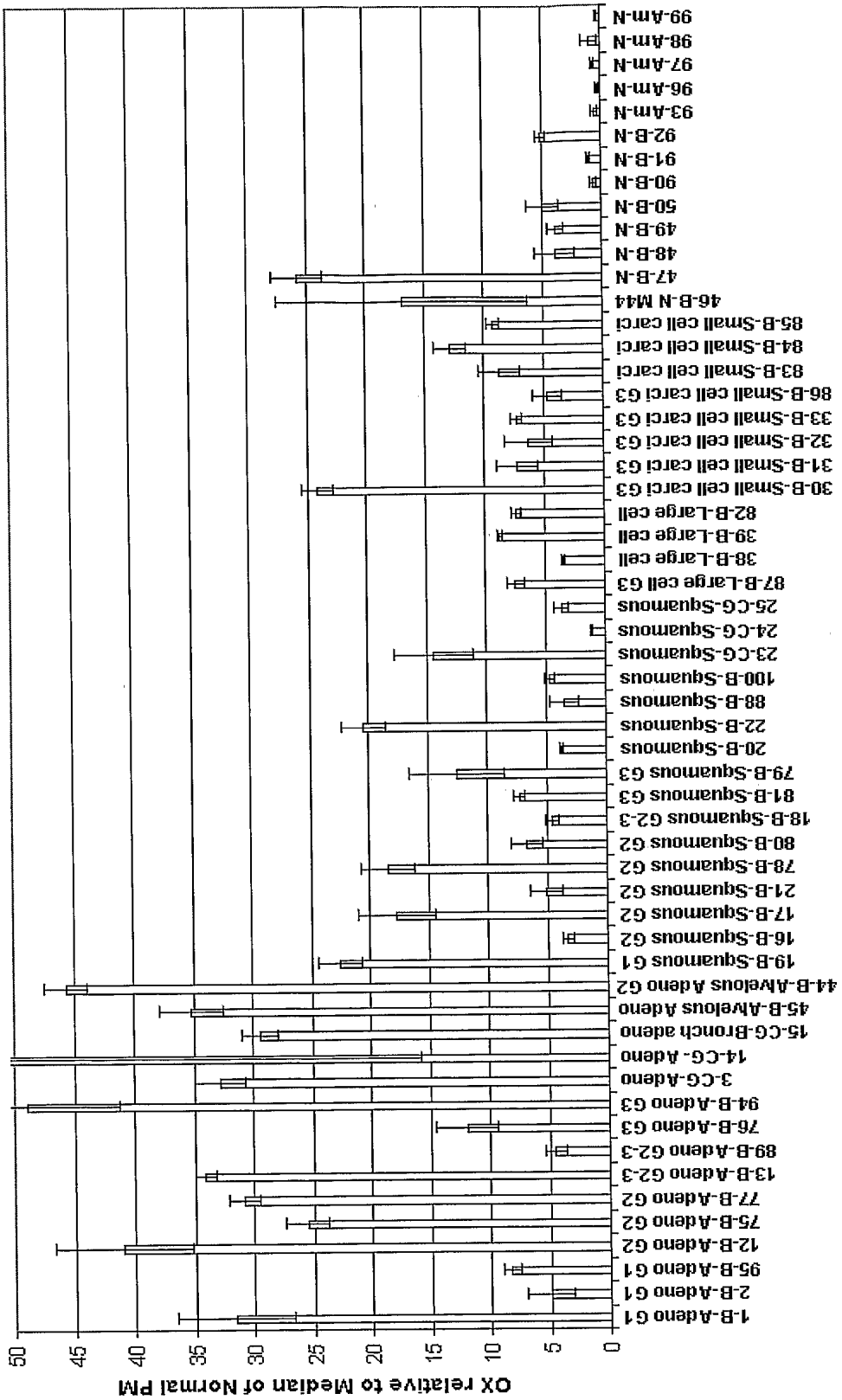

OX relative to Median of Normal PM
0
5
10
15
20
25
30
35
40
45
50
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 39

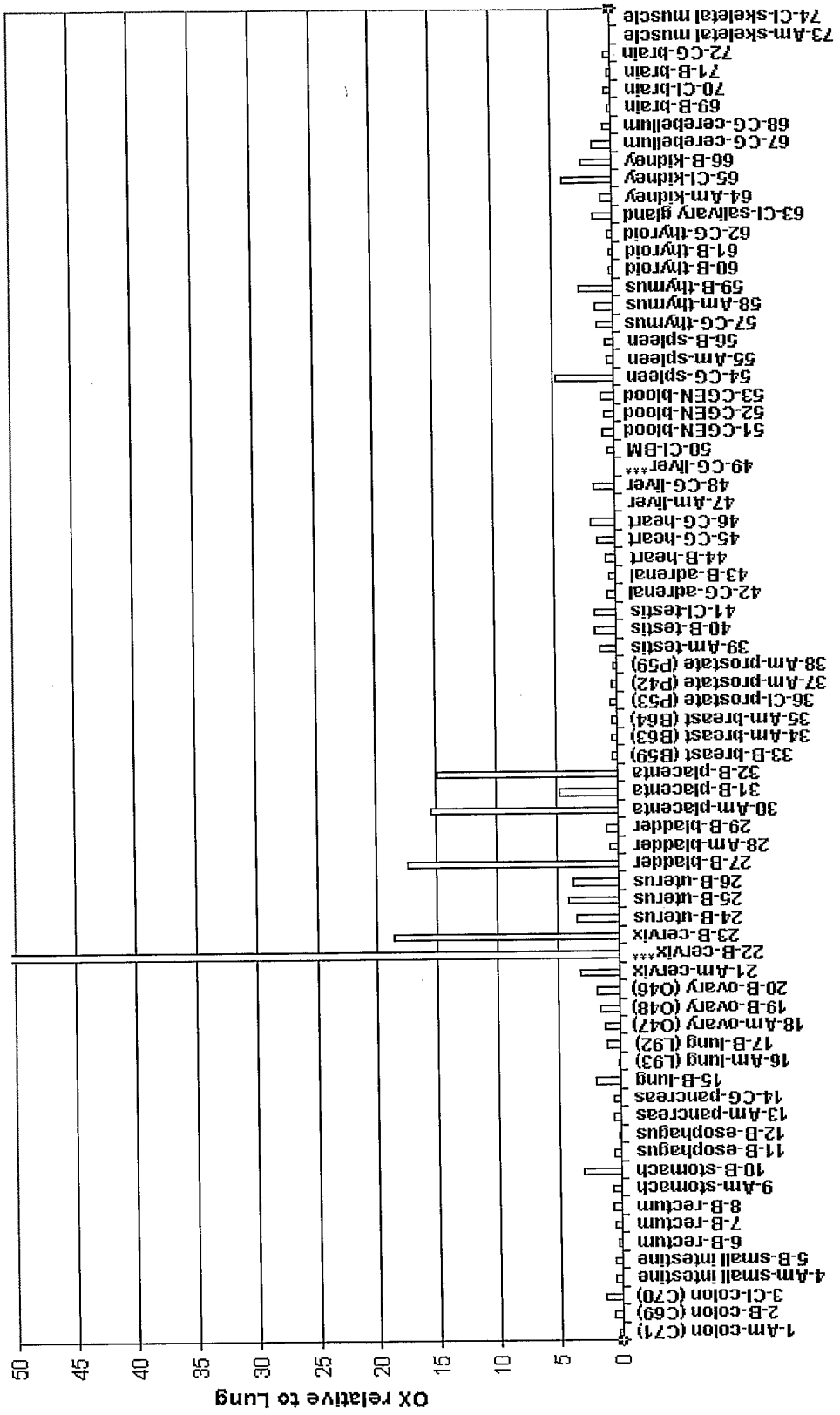

OX relative to Lung
0
5
10
15
20
25
30
35
40
45
50
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 40

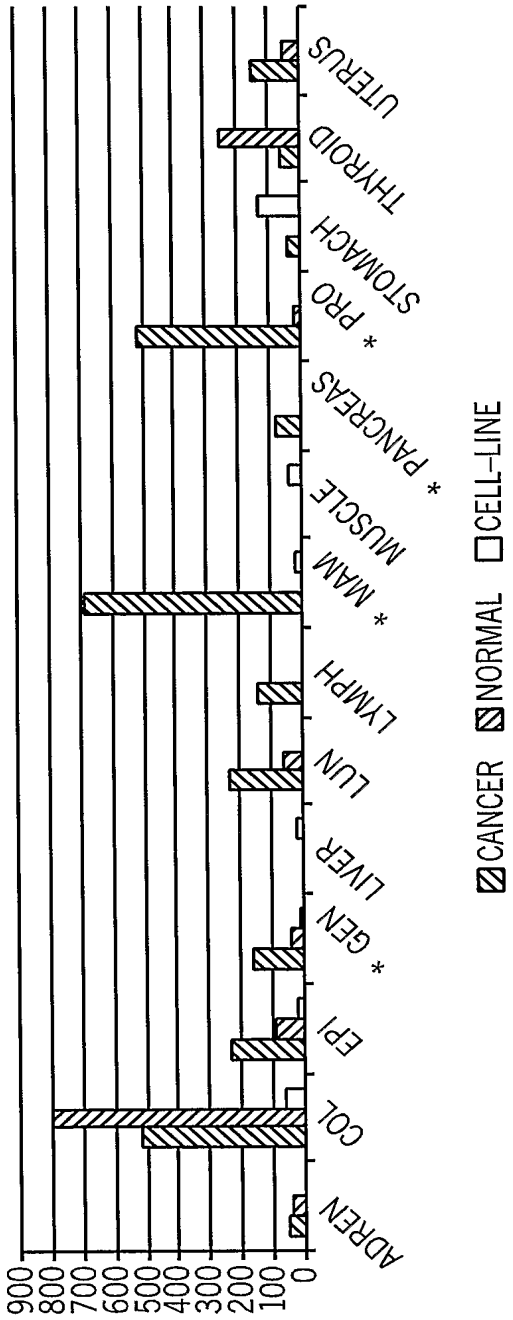

900
800
700
600
500
400
300
200
100
0
ADREN
COL
EPI
* GEN
LIVER
LUN
LYMPH
* MAM
MUSCLE
* PANCREAS
* PRO
STOMACH
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 41

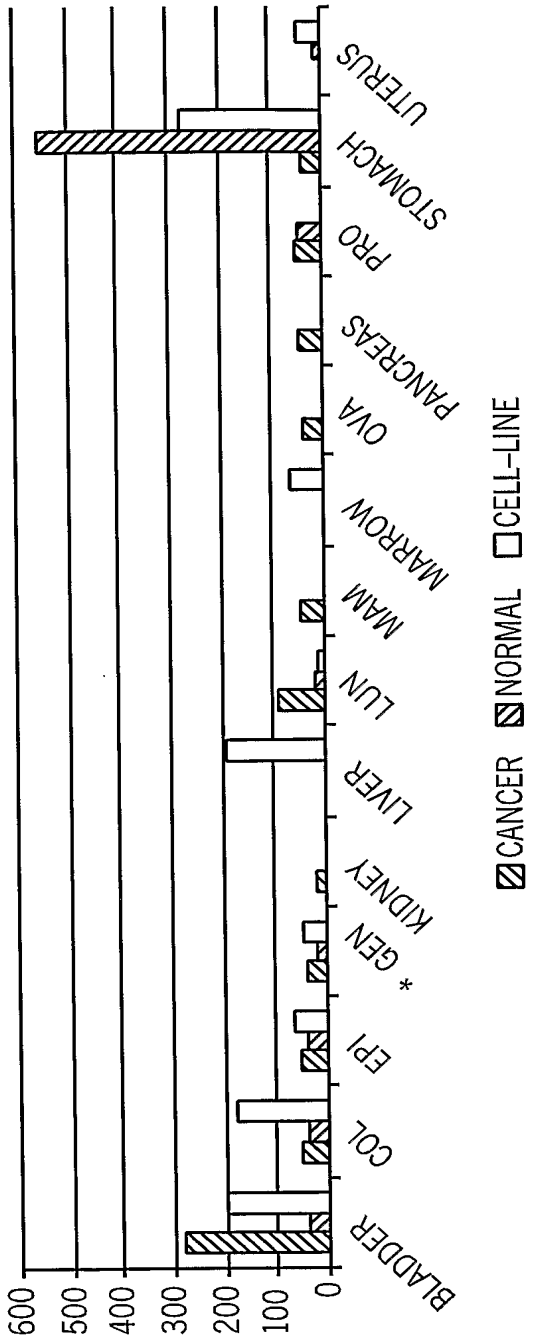

UTERUS
STOMACH
PRO
PANCREAS
OVA
MARROW
MAM
LUN
LIVER
KIDNEY
* GEN
EPI
COL
BLADDER
600
500
400
300
200
100
0
CANCER
NORMAL
CELL-LINE

FIG. 42

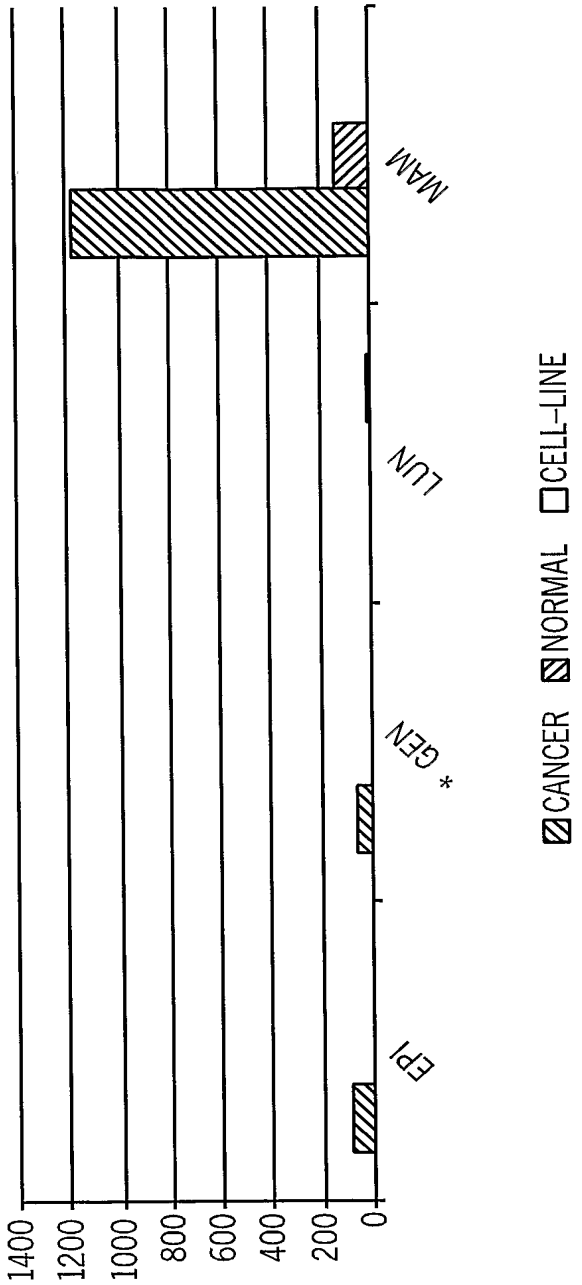

1400
1200
1000
800
600
400
200
0
EPI
GEN
LUN
MAM
*
CANCER
NORMAL
CELL-LINE

FIG. 43

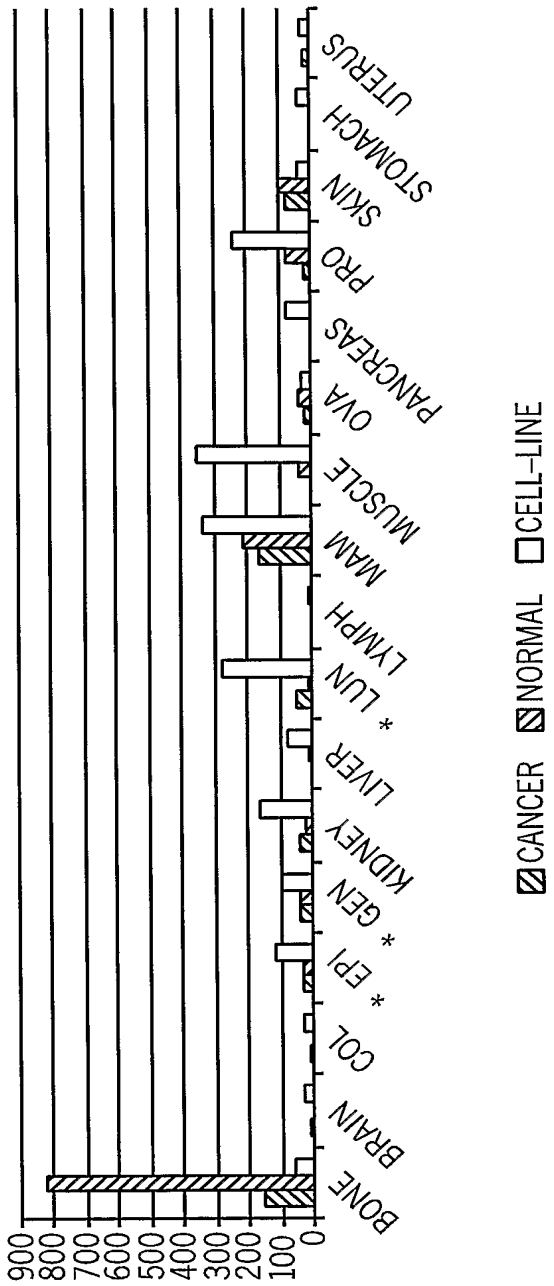

900
800
700
600
500
400
300
200
100
0
BONE
BRAIN
COL
* EPI
* GEN
KIDNEY
LIVER
* LUN
LYMPH
MAM
MUSCLE
OVA
PANCREAS
PRO
SKIN
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 44

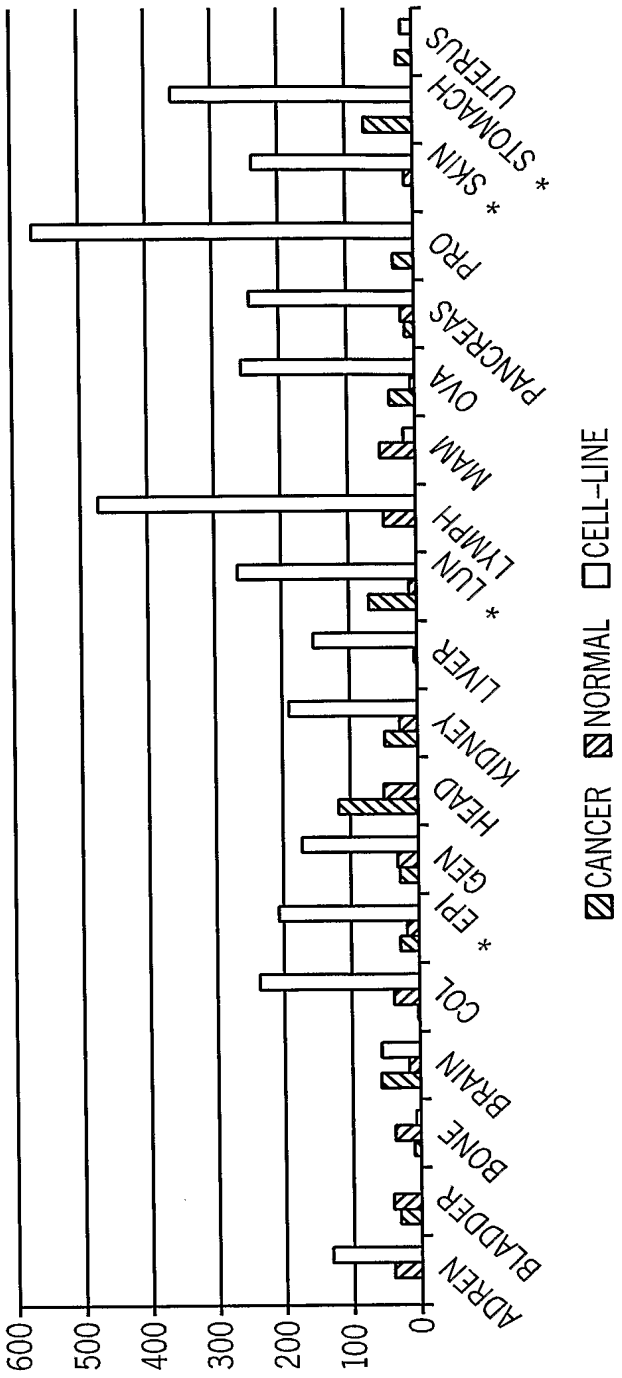

600
500
400
300
200
100
0
ADREN
BLADDER
BONE
BRAIN
COL
* EPI GEN
HEAD
KIDNEY
LIVER
* LUN
LYMPH
MAM
OVA
PANCREAS
PRO
* SKIN
* STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 45

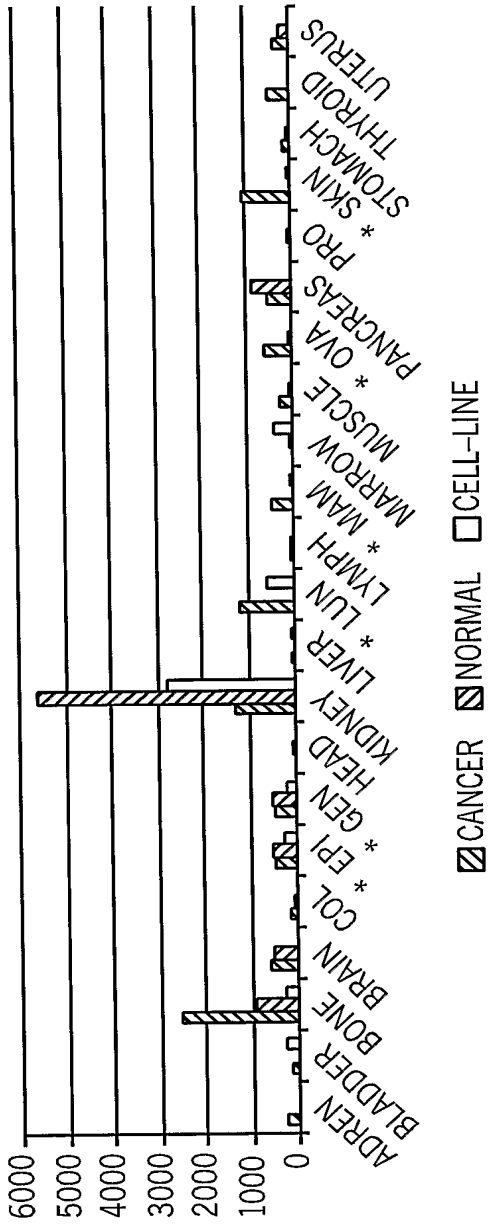

6000
5000
4000
3000
2000
1000
0
ADREN
BLADDER
BONE
BRAIN
COL
* EPI
* GEN
HEAD
KIDNEY
LIVER
* LUN
LYMPH
* MAM
MARROW
MUSCLE
* OVA
PANCREAS
PRO
* SKIN
STOMACH
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 46

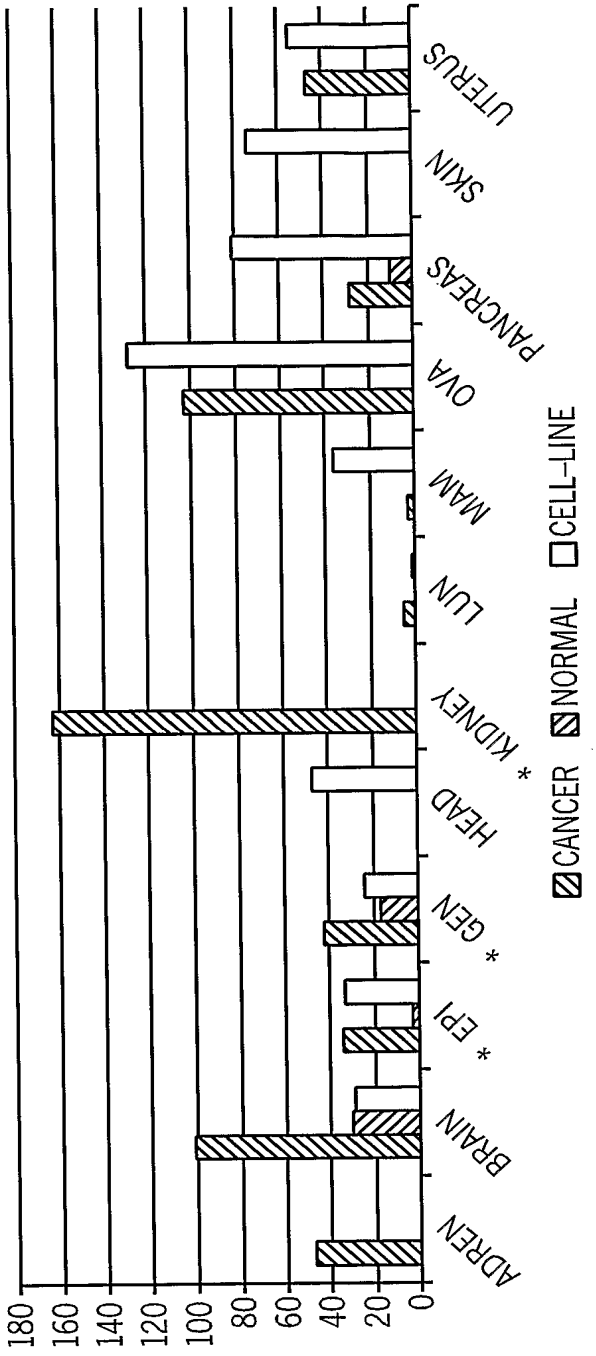

180
160
140
120
100
80
60
40
20
0
ADREN
BRAIN
* EPI
* GEN
HEAD
* KIDNEY
LUN
MAM
OVA
PANCREAS
SKIN
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 47

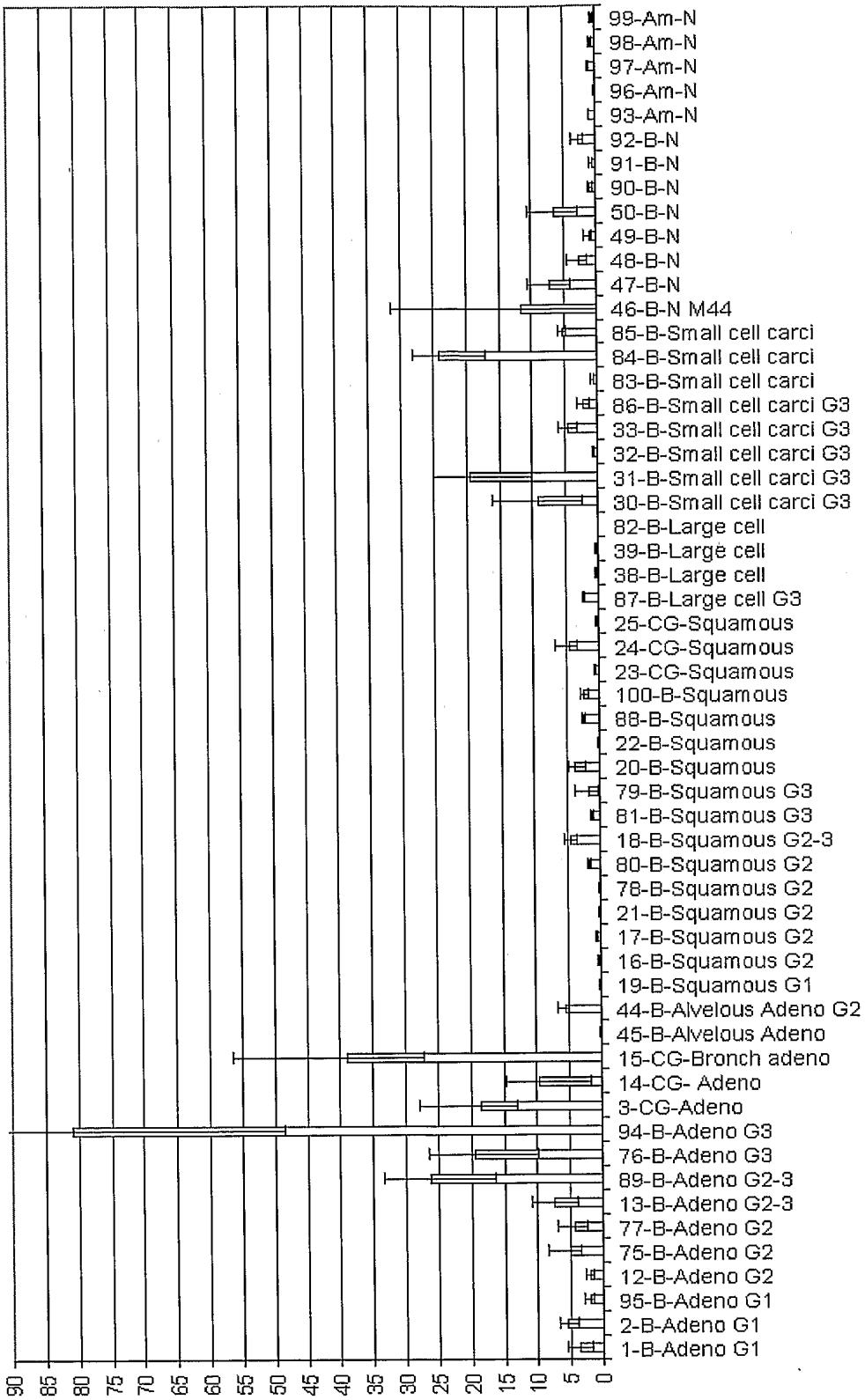

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
90 85 80 75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0

FIG. 48

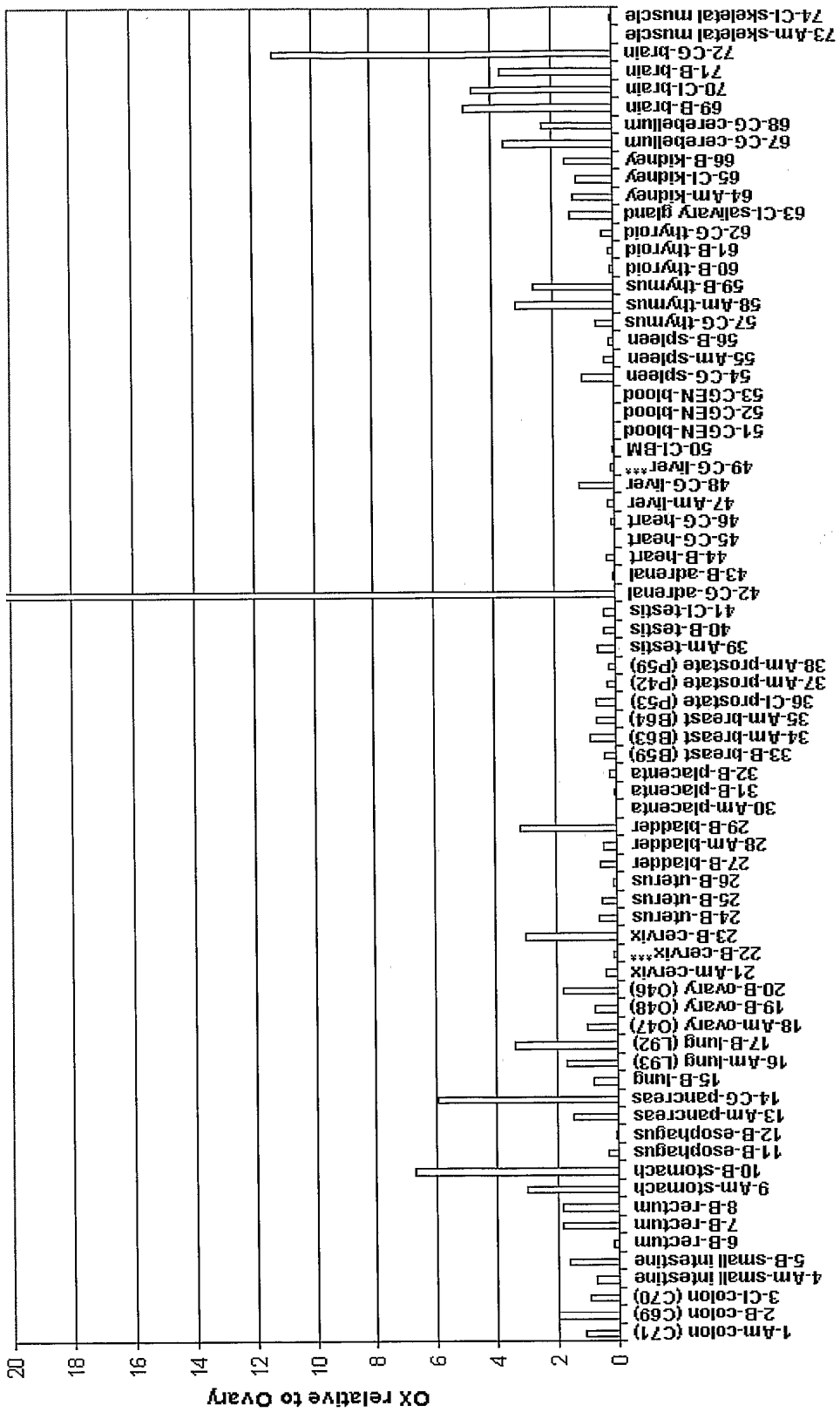

OX relative to Ovary
0
2
4
6
8
10
12
14
16
18
20
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 49

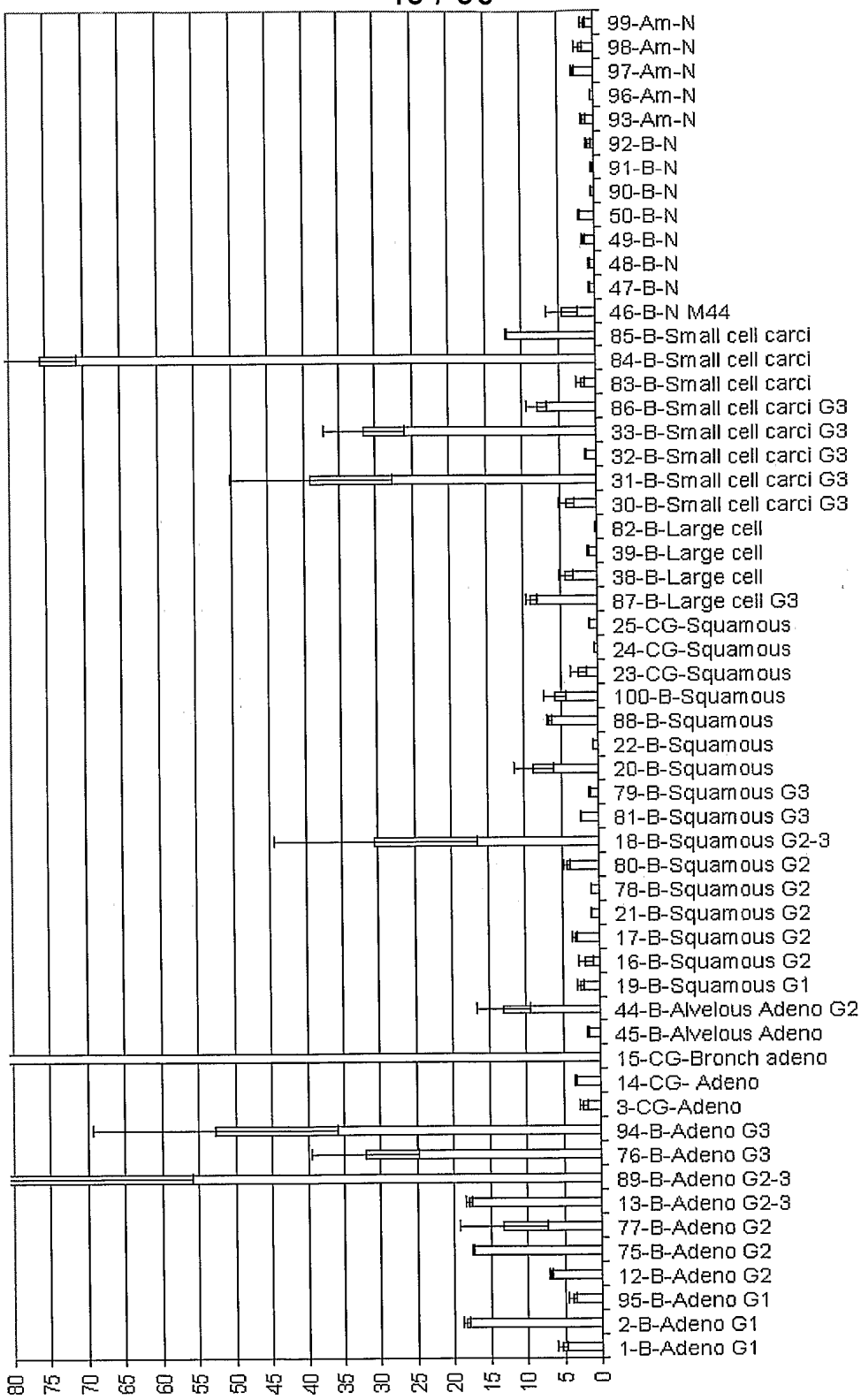

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
80 75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0

FIG. 50

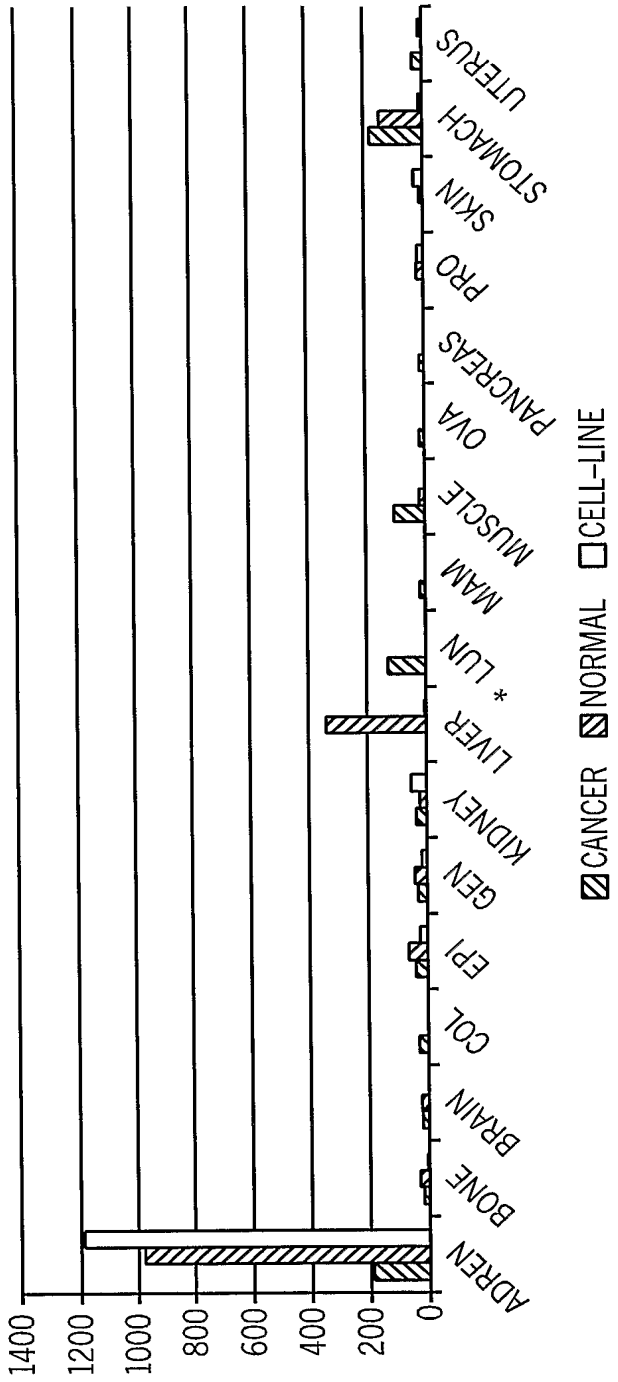

1400
1200
1000
800
600
400
200
0
ADREN
BONE
BRAIN
COL
EPI
GEN
KIDNEY
LIVER
* LUN
MAM
MUSCLE
OVA
PANCREAS
PRO
SKIN
STOMACH
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 51

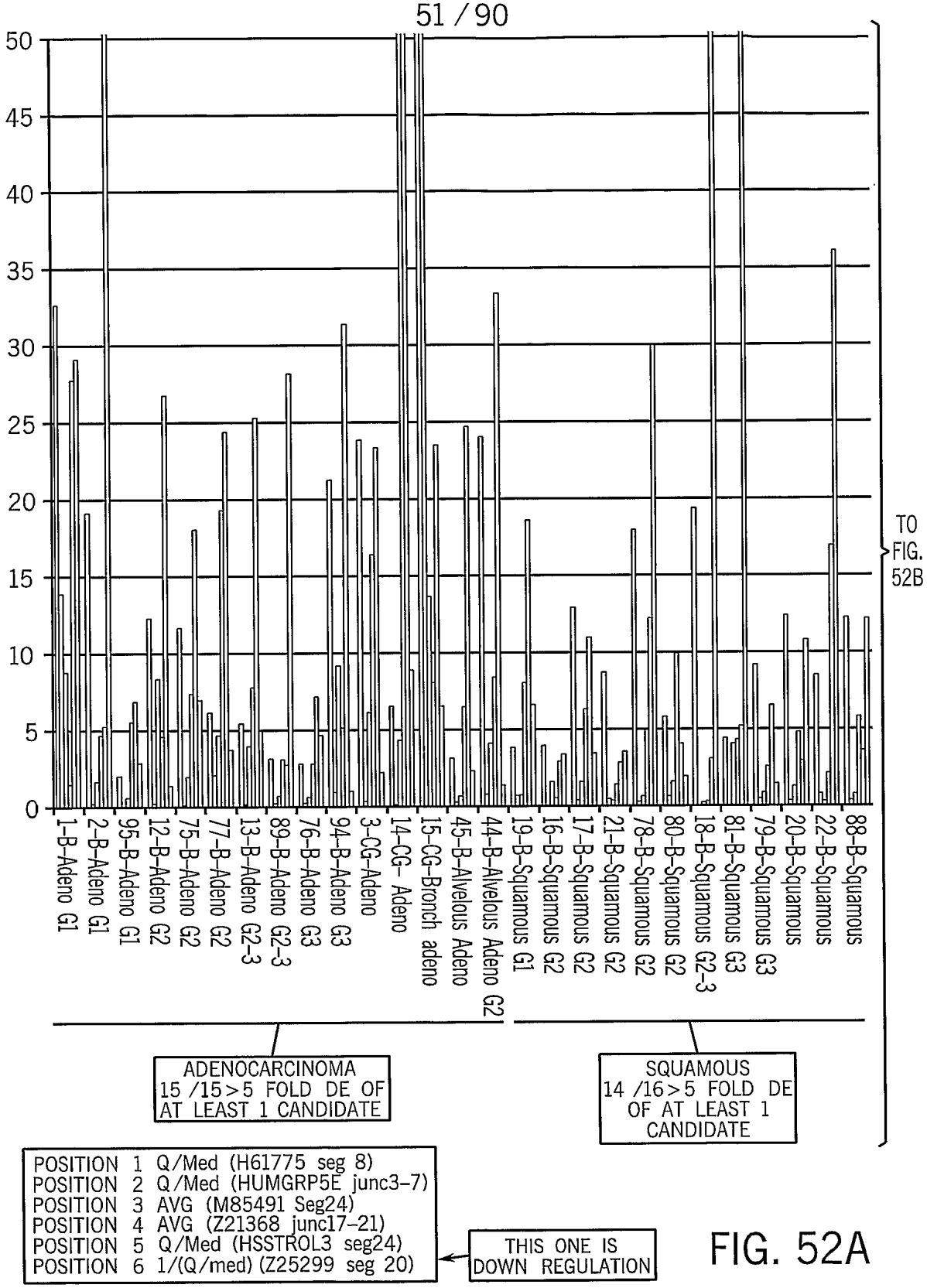

50
45
40
35
30
25
20
15
10
5
0
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alvelous Adeno
44-B-Alvelous Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
TO FIG. 52B
ADENOCARCINOMA 15 /15>5 FOLD DE OF AT LEAST 1 CANDIDATE
SQUAMOUS 14 /16>5 FOLD DE OF AT LEAST 1 CANDIDATE
POSITION 1 Q/Med (H61775 seg 8)
POSITION 2 Q/Med (HUMGRP5E junc3-7)
POSITION 3 AVG (M85491 Seg24)
POSITION 4 AVG (Z21368 junc17-21)
POSITION 5 Q/Med (HSSTROL3 seg24)
POSITION 6 1/(Q/med) (Z25299 seg 20)
THIS ONE IS DOWN REGULATION

FIG. 52A

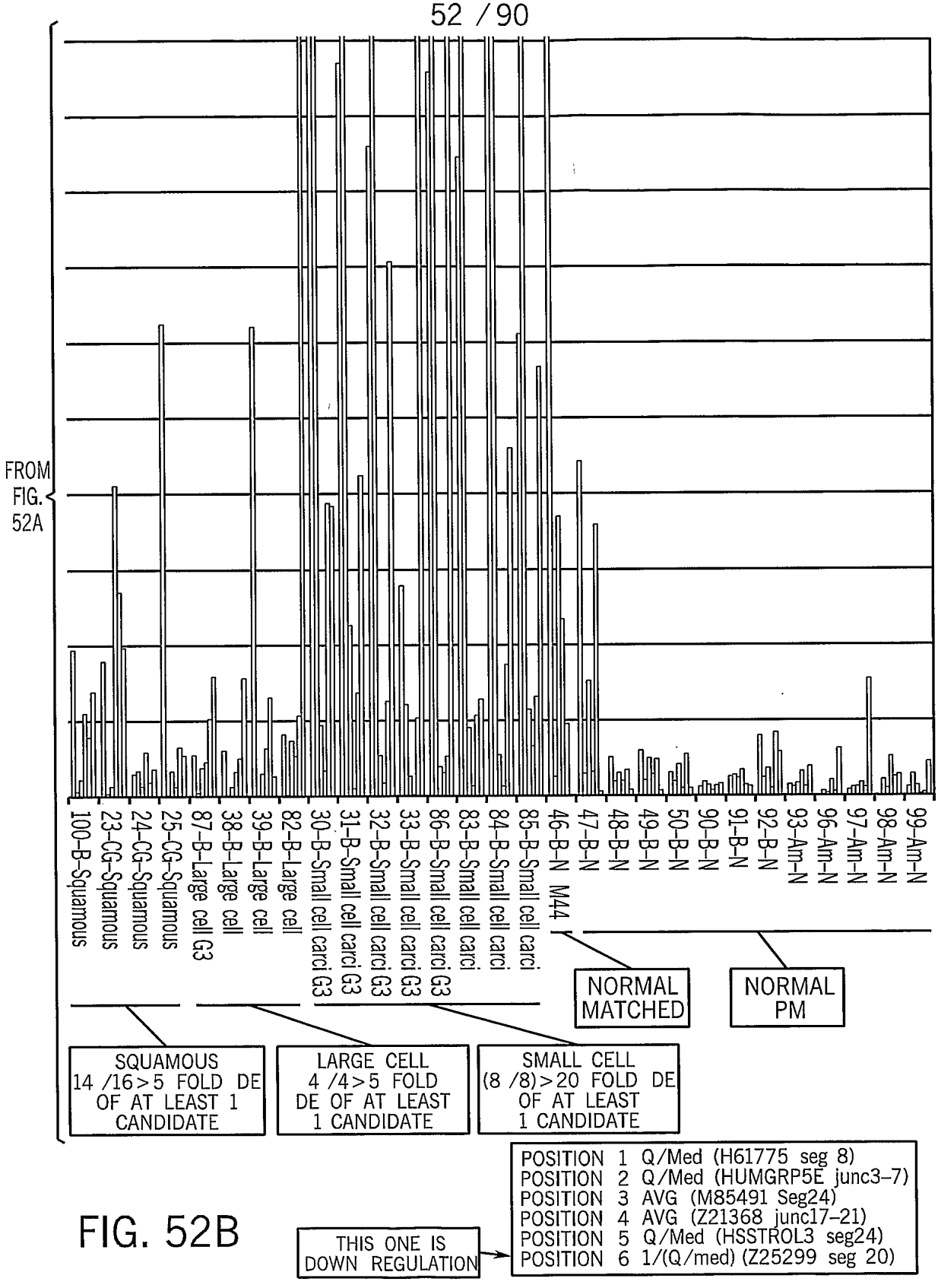

FROM FIG. 52A
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N
NORMAL MATCHED
NORMAL PM
SQUAMOUS 14 /16>5 FOLD DE OF AT LEAST 1 CANDIDATE
LARGE CELL 4 /4>5 FOLD DE OF AT LEAST 1 CANDIDATE
SMALL CELL (8 /8)>20 FOLD DE OF AT LEAST 1 CANDIDATE
POSITION 1 Q/Med (H61775 seg 8)
POSITION 2 Q/Med (HUMGRP5E junc3-7)
POSITION 3 AVG (M85491 Seg24)
POSITION 4 AVG (Z21368 junc17-21)
POSITION 5 Q/Med (HSSTROL3 seg24)
POSITION 6 1/(Q/med) (Z25299 seg 20)
THIS ONE IS DOWN REGULATION

FIG. 52B

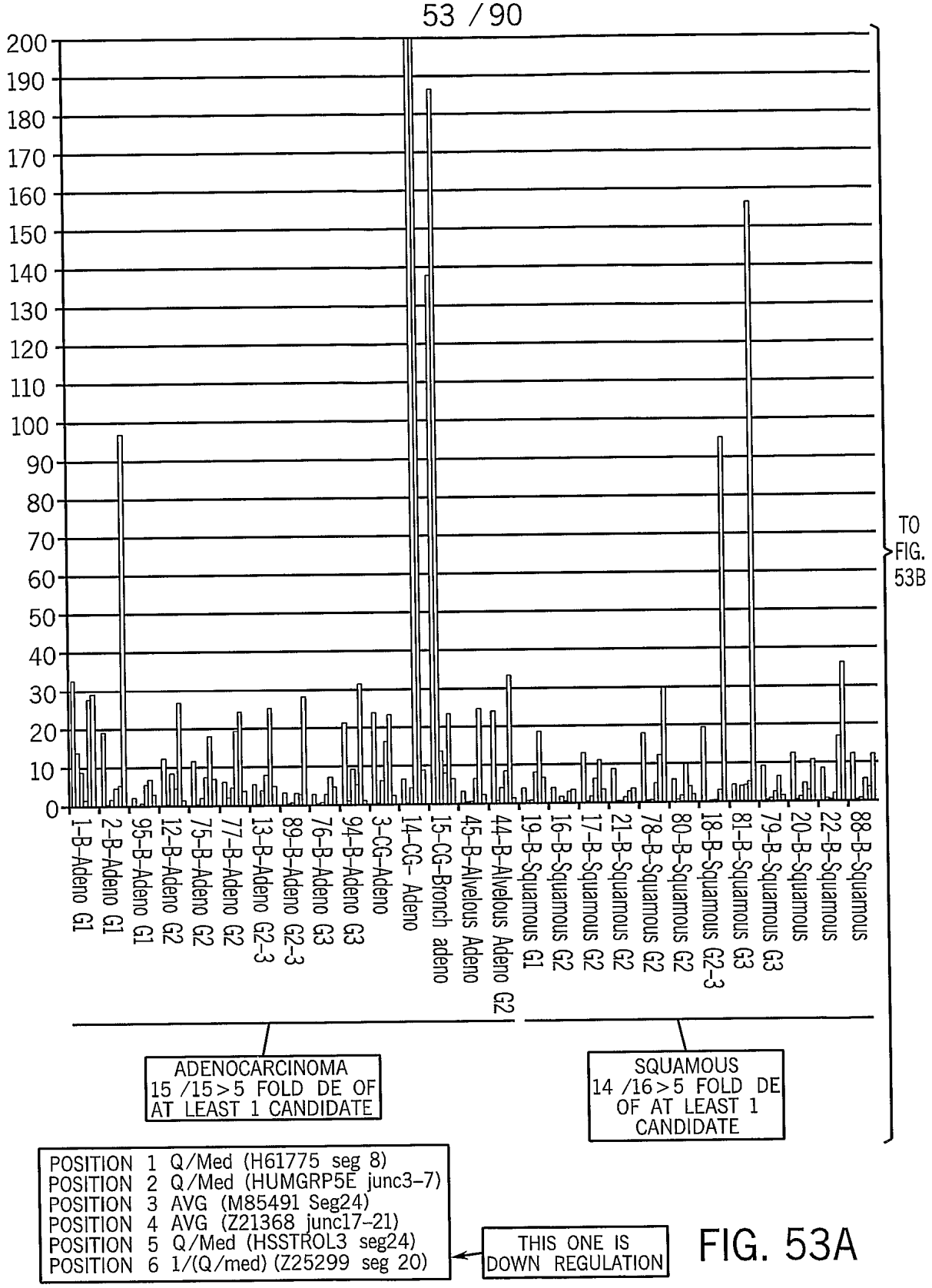

200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alvelous Adeno
44-B-Alvelous Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
TO FIG. 53B
ADENOCARCINOMA
15 /15>5 FOLD DE OF AT LEAST 1 CANDIDATE
SQUAMOUS
14 /16>5 FOLD DE OF AT LEAST 1 CANDIDATE
POSITION 1 Q/Med (H61775 seg 8)
POSITION 2 Q/Med (HUMGRP5E junc3-7)
POSITION 3 AVG (M85491 Seg24)
POSITION 4 AVG (Z21368 junc17-21)
POSITION 5 Q/Med (HSSTROL3 seg24)
POSITION 6 1/(Q/med) (Z25299 seg 20)
THIS ONE IS DOWN REGULATION

FIG. 53A

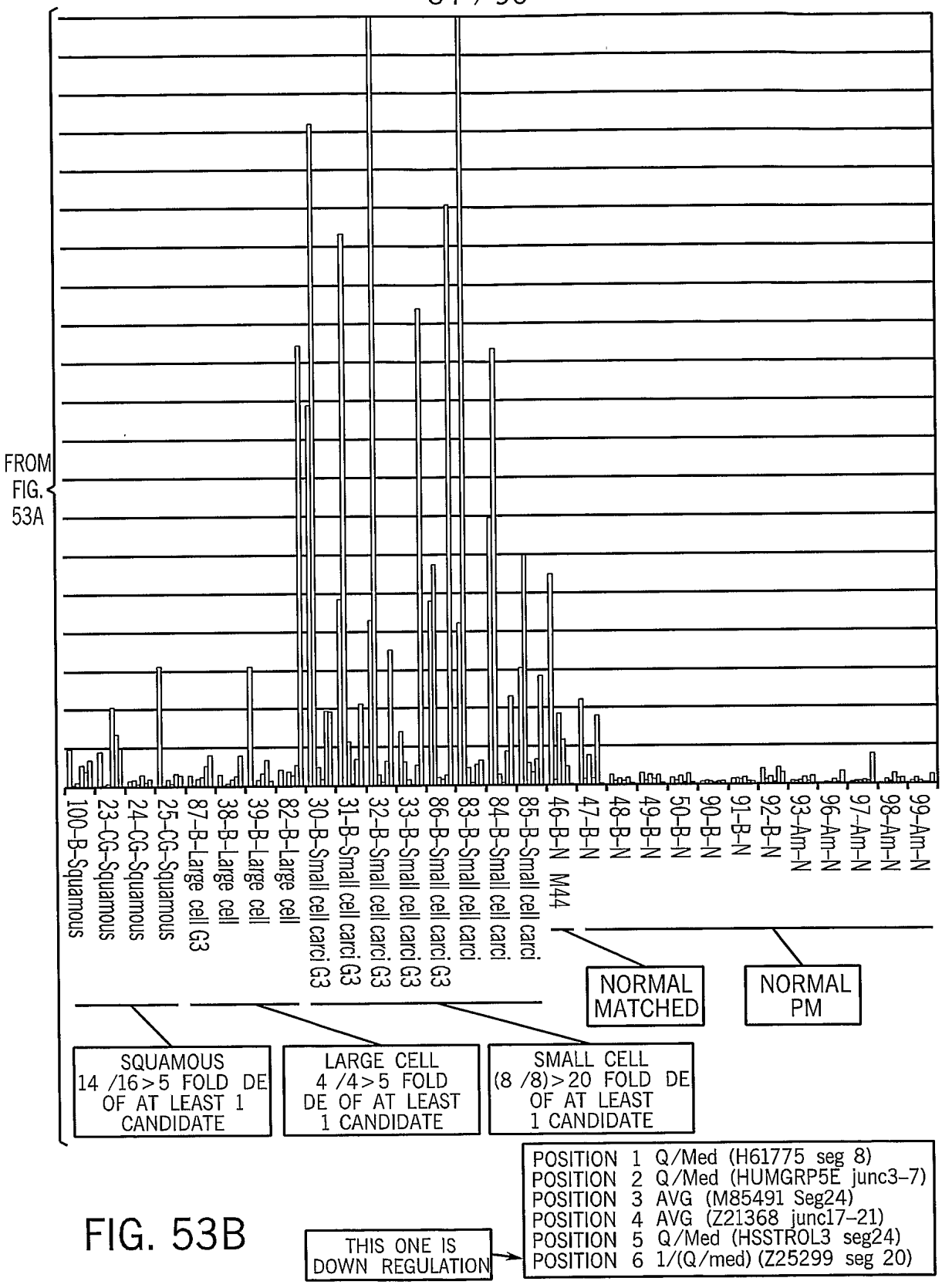

FROM FIG. 53A
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N
NORMAL MATCHED
NORMAL PM
SQUAMOUS 14 /16>5 FOLD DE OF AT LEAST 1 CANDIDATE
LARGE CELL 4 /4>5 FOLD DE OF AT LEAST 1 CANDIDATE
SMALL CELL (8 /8)>20 FOLD DE OF AT LEAST 1 CANDIDATE
POSITION 1 Q/Med (H61775 seg 8)
POSITION 2 Q/Med (HUMGRP5E junc3-7)
POSITION 3 AVG (M85491 Seg24)
POSITION 4 AVG (Z21368 junc17-21)
POSITION 5 Q/Med (HSSTROL3 seg24)
POSITION 6 1/(Q/med) (Z25299 seg 20)
THIS ONE IS DOWN REGULATION

FIG. 53B

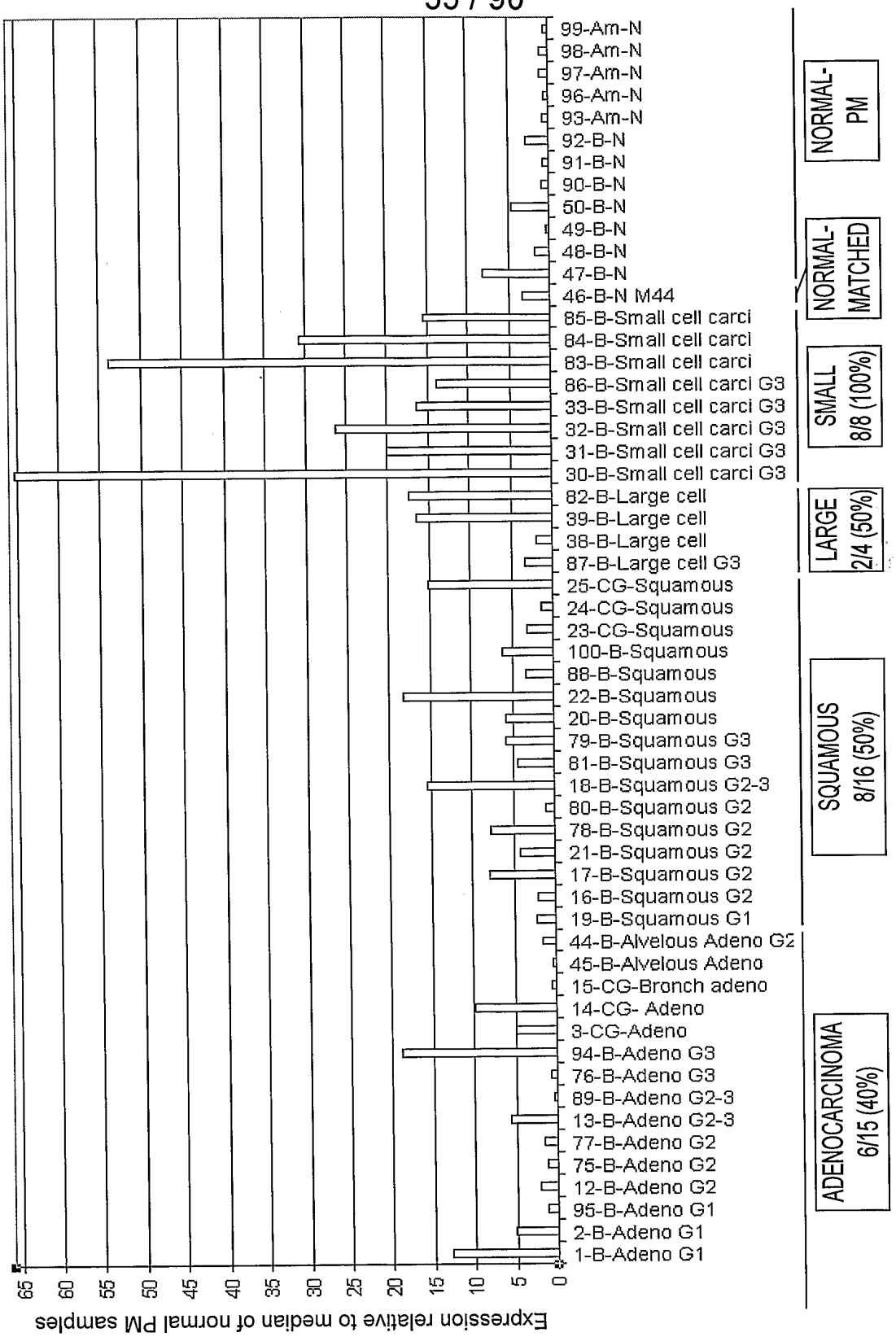

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 2/4 (50%)
SQUAMOUS 8/16 (50%)
ADENOCARCINOMA 6/15 (40%)
0
5
10
15
20
25
30
35
40
45
50
55
60
65
Expression relative to median of normal PM samples

FIG. 54A

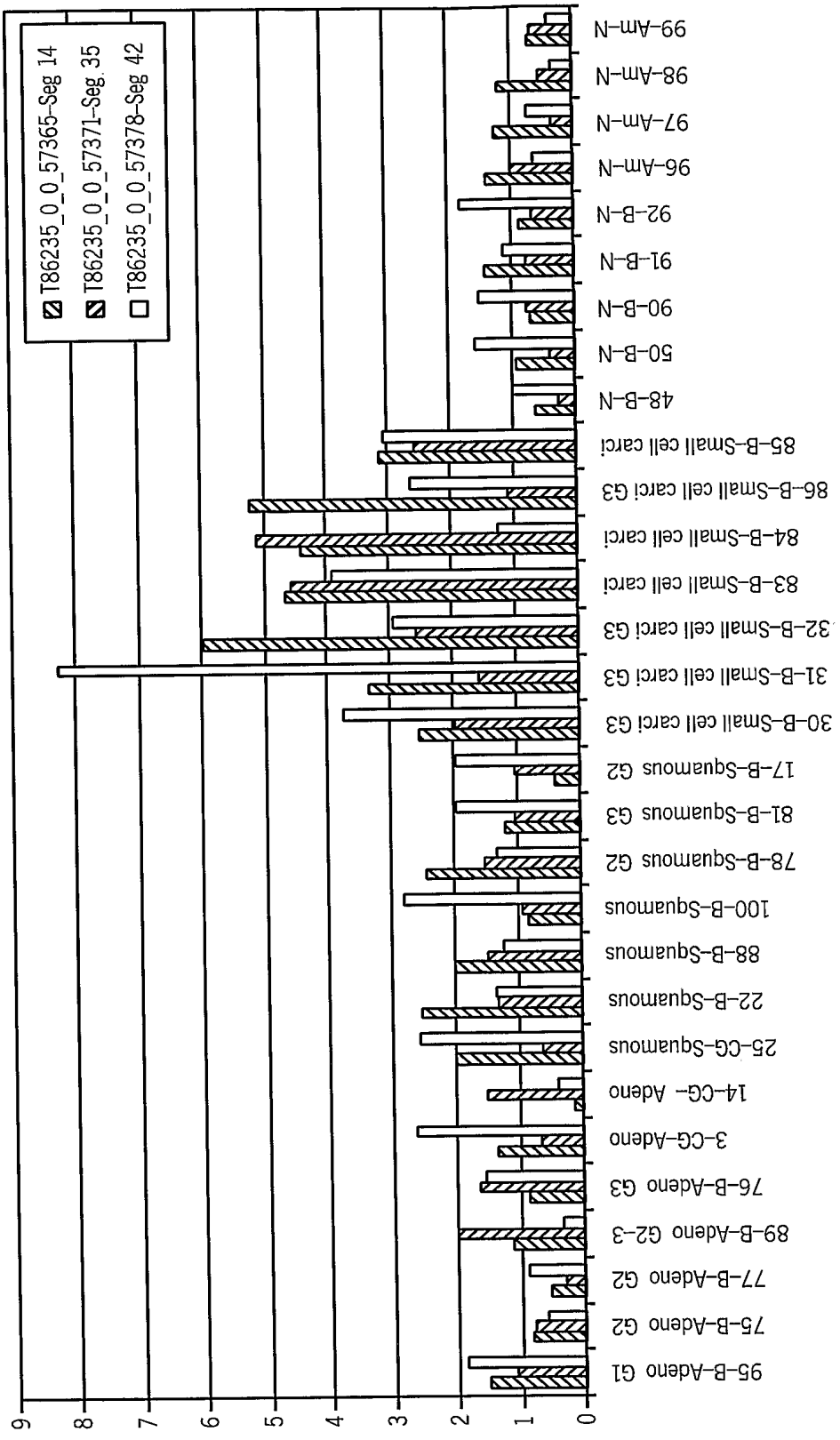

T86235_0_0_57365-Seg 14
T86235_0_0_57371-Seg 35
T86235_0_0_57378-Seg 42
0
1
2
3
4
5
6
7
8
9
95-B-Adeno G1
75-B-Adeno G2
77-B-Adeno G2
89-B-Adeno G2-3
76-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
25-CG-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
78-B-Squamous G2
81-B-Squamous G3
17-B-Squamous G2
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
86-B-Small cell carci G3
85-B-Small cell carci
48-B-N
50-B-N
90-B-N
91-B-N
92-B-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 54B

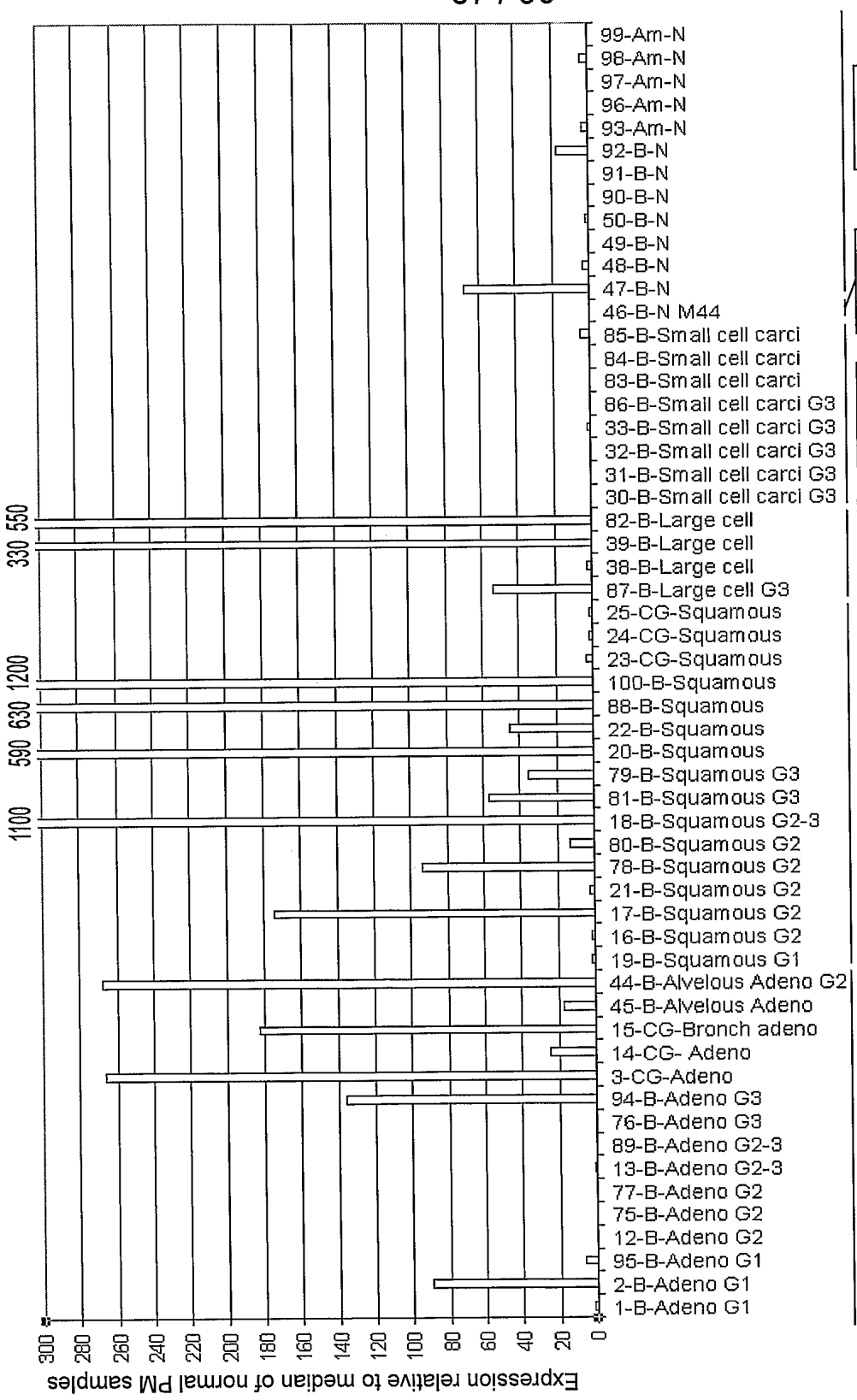

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 0/8
LARGE 3/4 (75%)
SQUAMOUS 9/16 (56%)
ADENOCARCINOMA 6/15 (40%)
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300
1100 590 630 1200 330 550
Expression relative to median of normal PM samples

FIG. 55

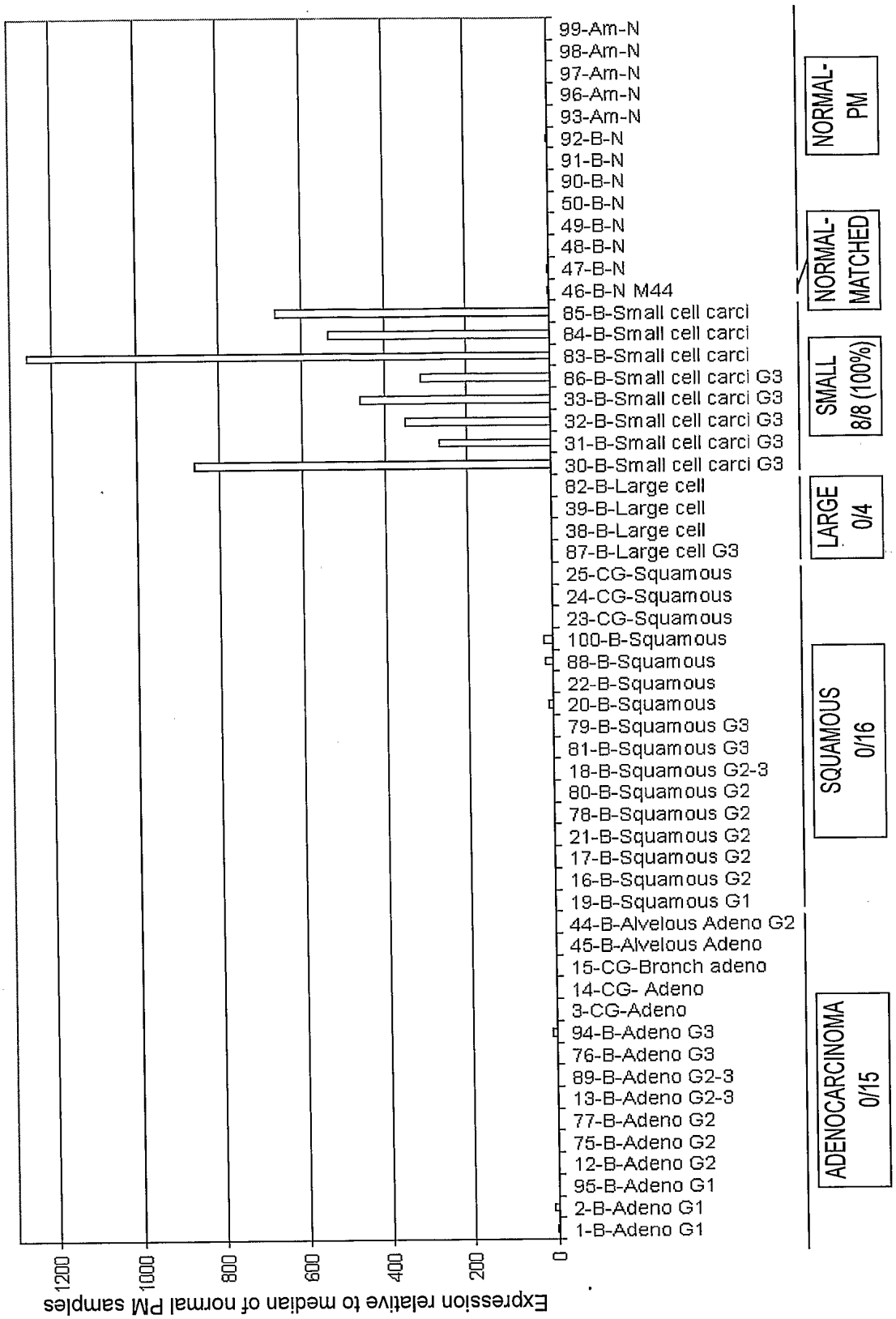

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 0/4
SQUAMOUS 0/16
ADENOCARCINOMA 0/15
0
200
400
600
800
1000
1200
Expression relative to median of normal PM samples

FIG. 56A

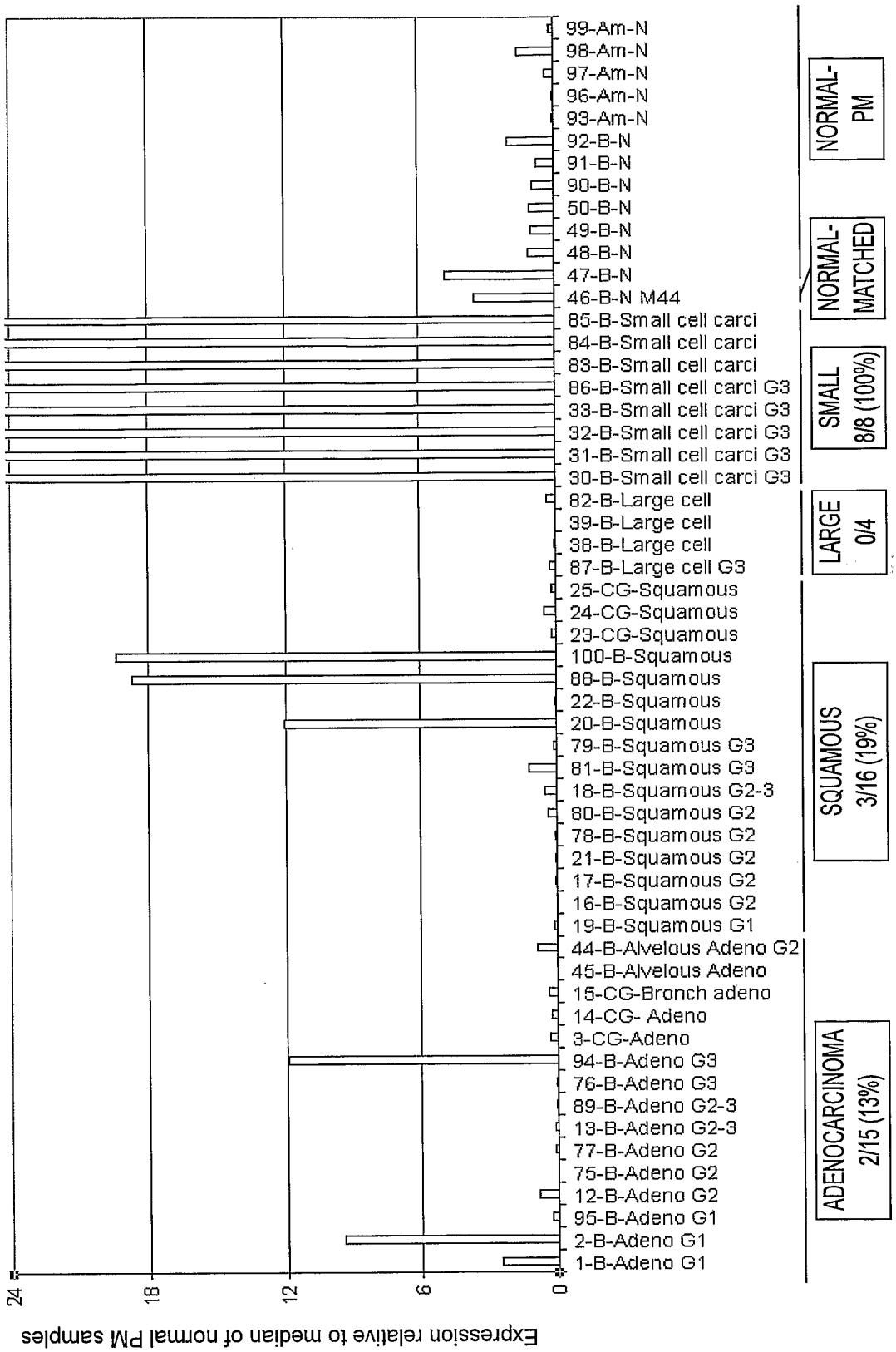

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 0/4
SQUAMOUS 3/16 (19%)
ADENOCARCINOMA 2/15 (13%)
0
6
12
18
24
Expression relative to median of normal PM samples

FIG. 56B

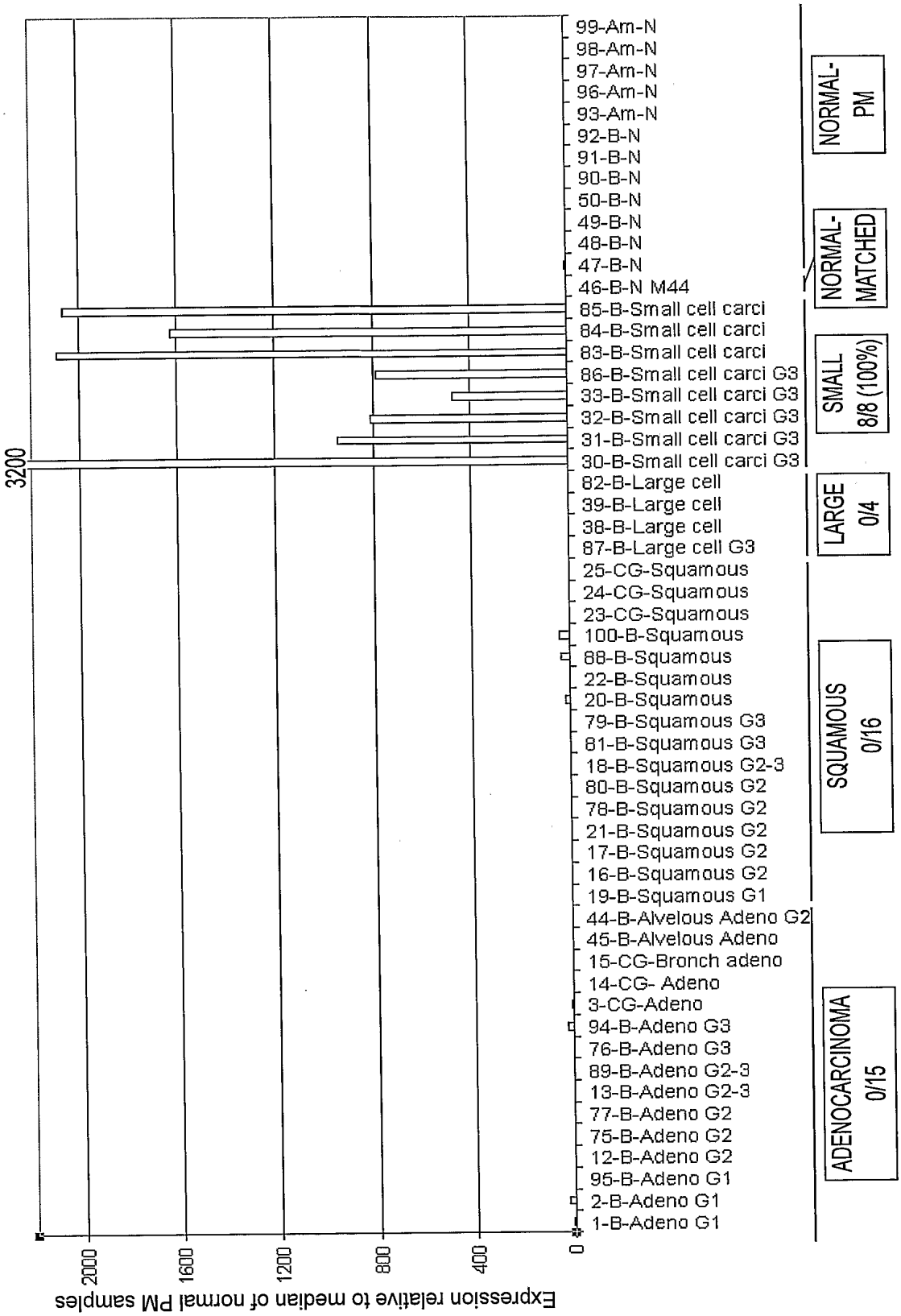

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 0/4
SQUAMOUS 0/16
ADENOCARCINOMA 0/15
3200
0
400
800
1200
1600
2000
Expression relative to median of normal PM samples

FIG. 57A

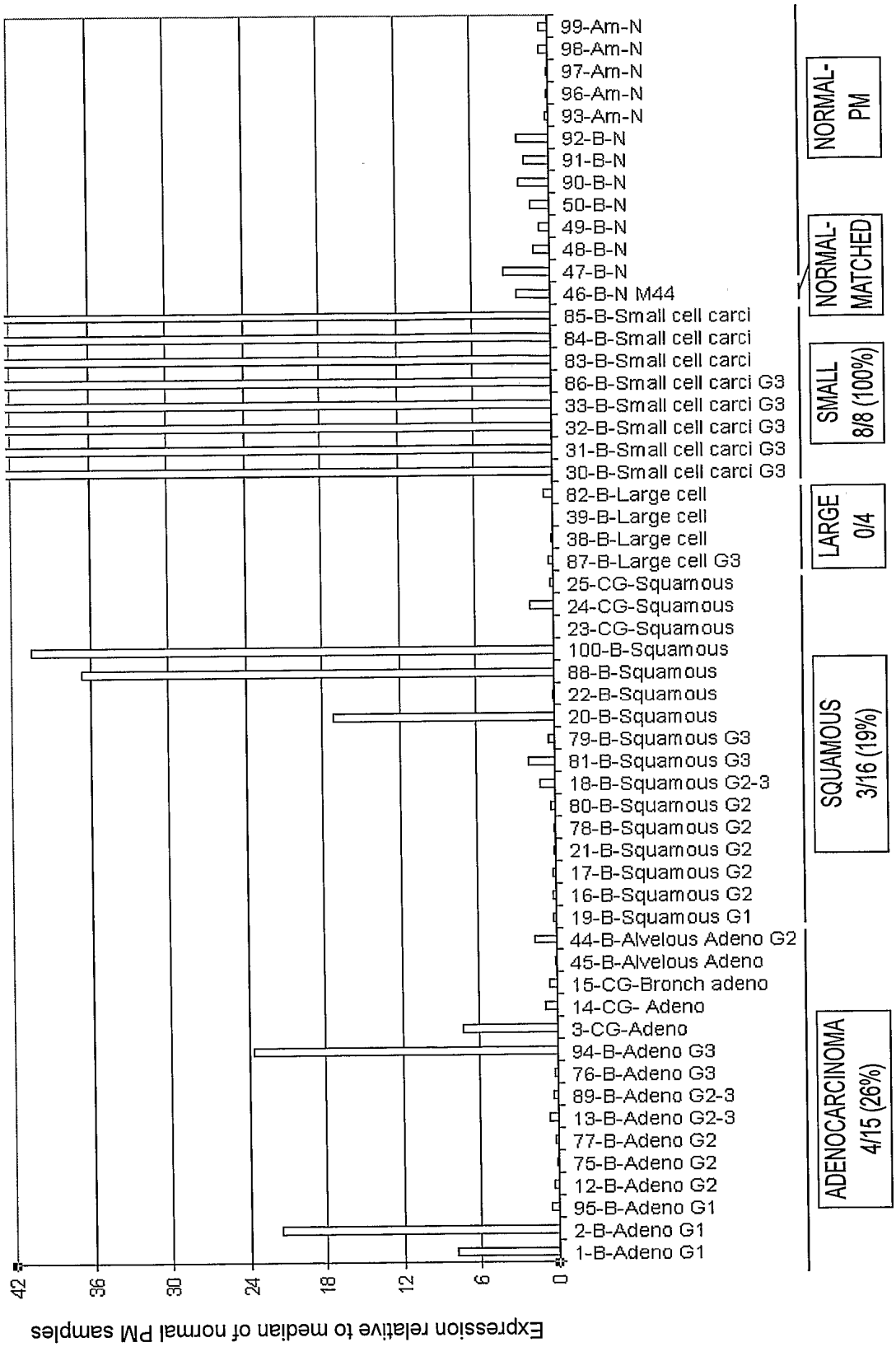

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 0/4
SQUAMOUS 3/16 (19%)
ADENOCARCINOMA 4/15 (26%)
42
36
30
24
18
12
6
0
Expression relative to median of normal PM samples

FIG. 57B

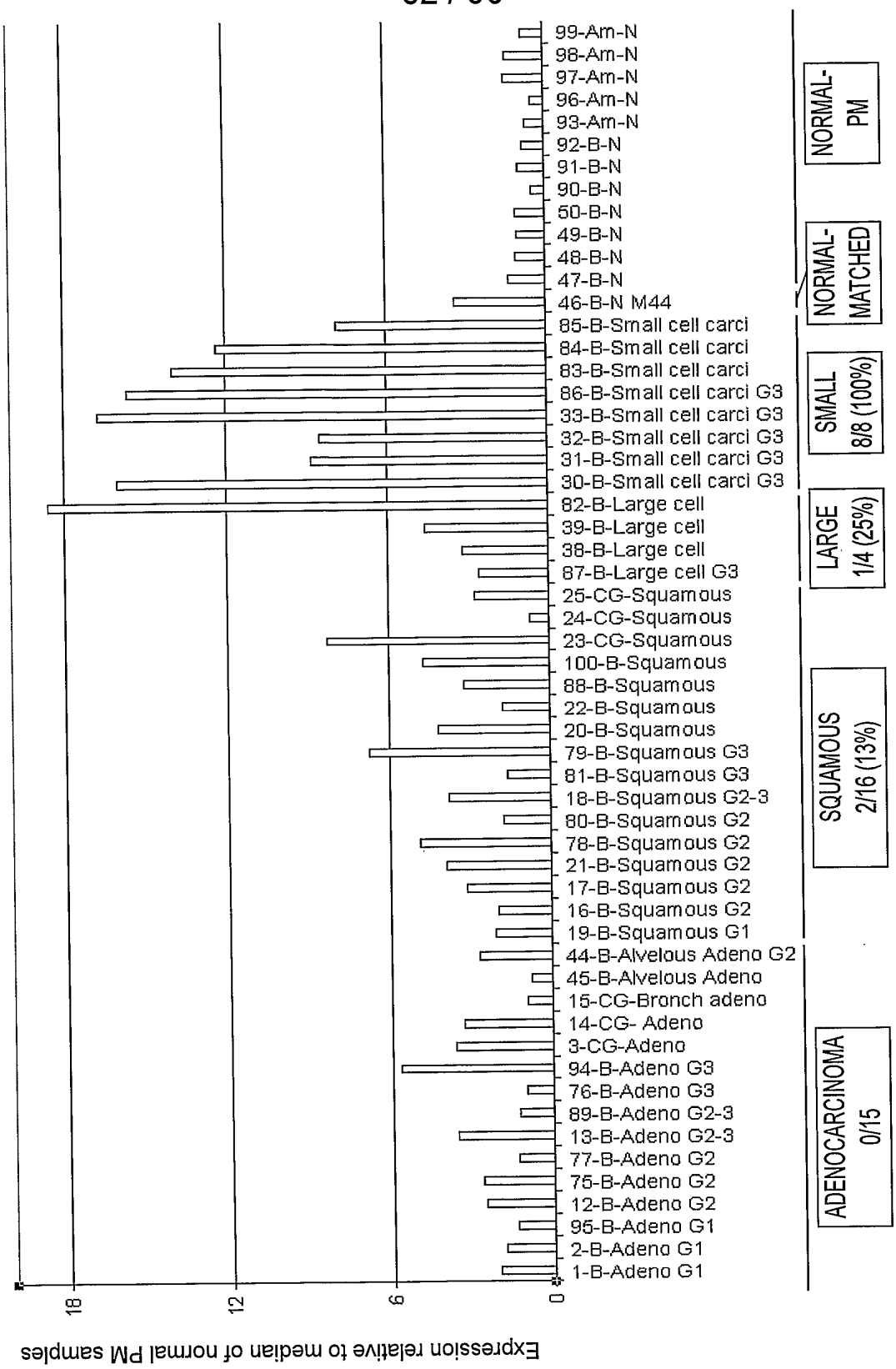

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 8/8 (100%)
LARGE 1/4 (25%)
SQUAMOUS 2/16 (13%)
ADENOCARCINOMA 0/15
0
6
12
18
Expression relative to median of normal PM samples

FIG. 58

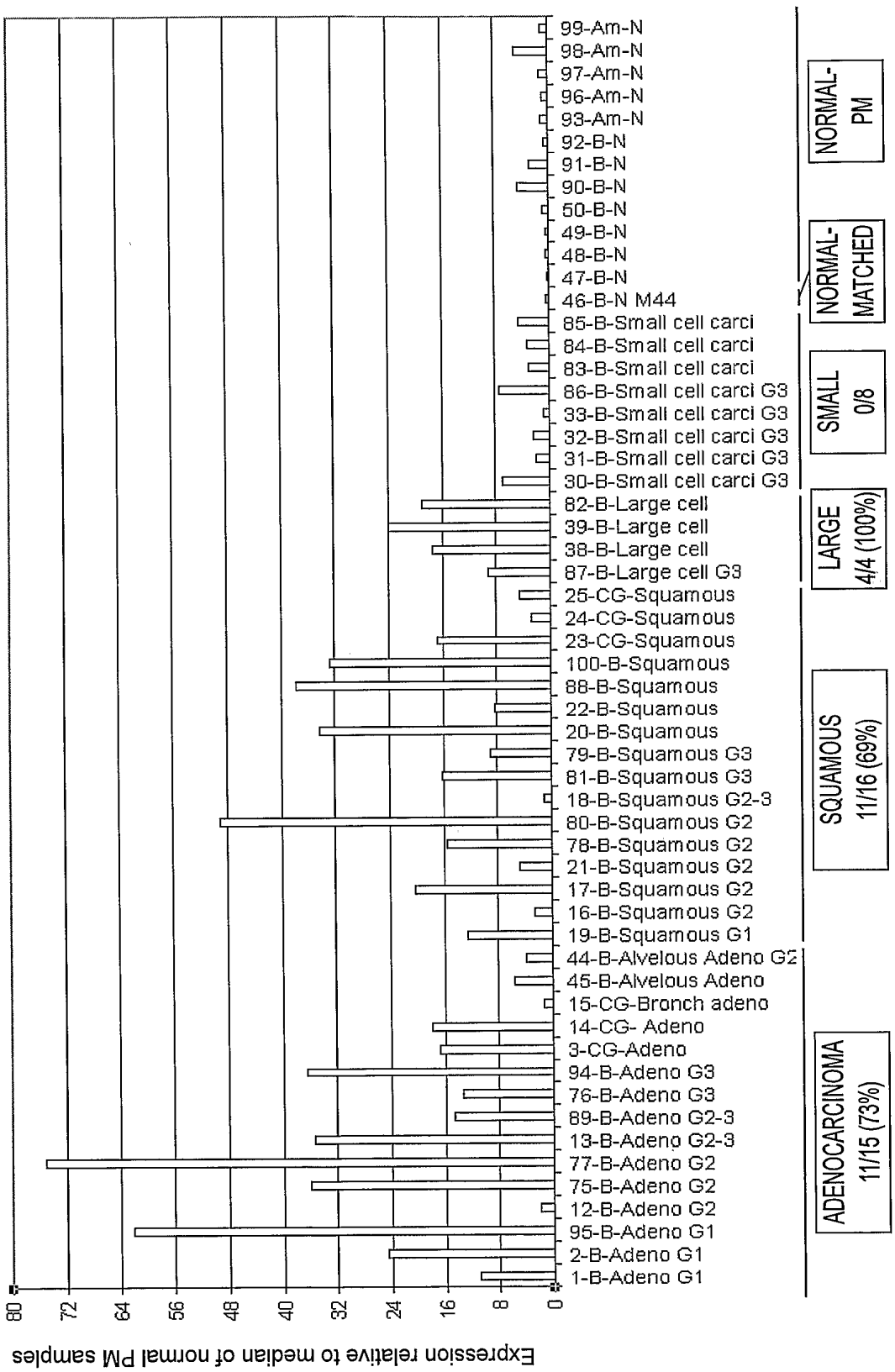

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 0/8
LARGE 4/4 (100%)
SQUAMOUS 11/16 (69%)
ADENOCARCINOMA 11/15 (73%)
80
72
64
56
48
40
32
24
16
8
0
Expression relative to median of normal PM samples

FIG. 59

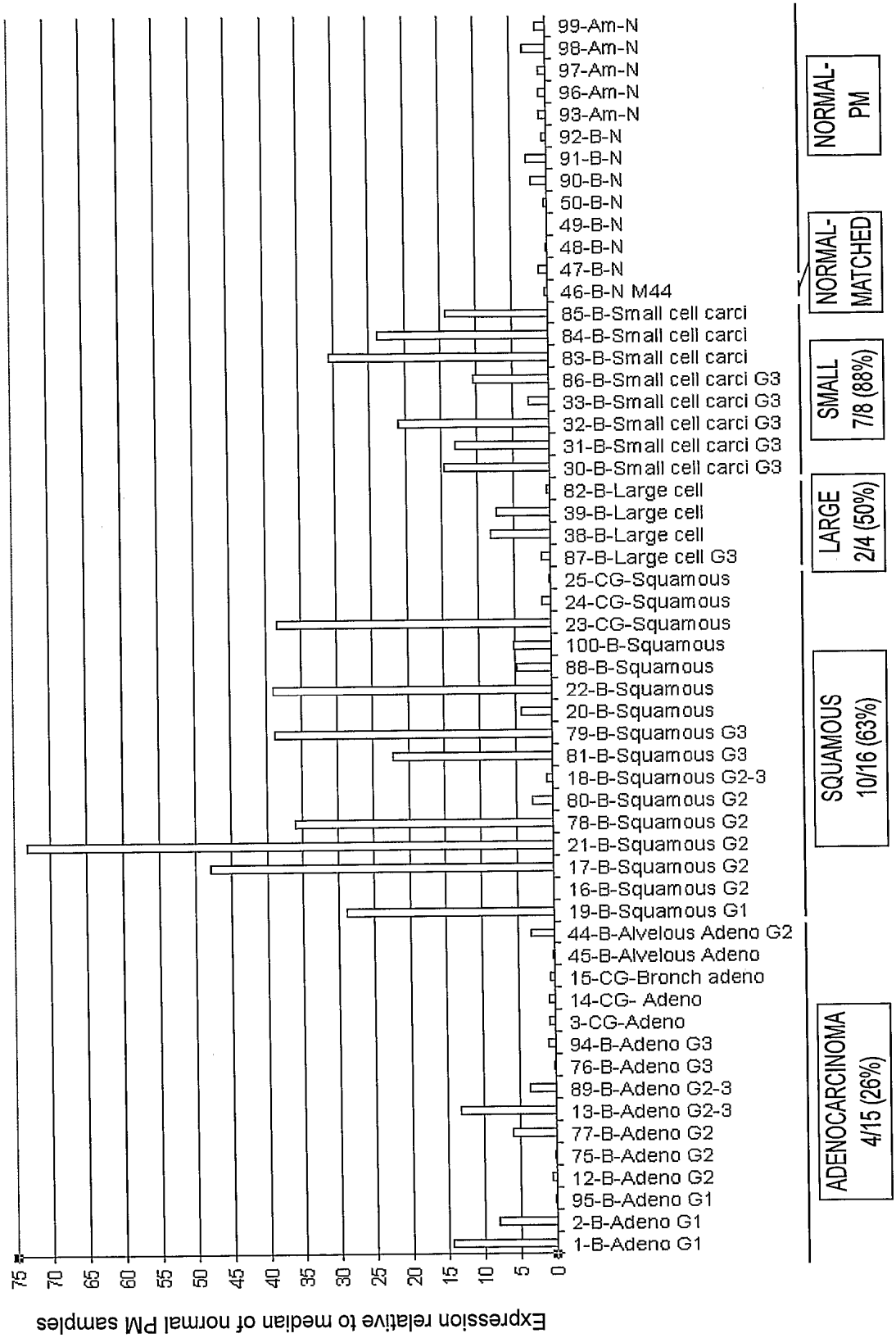

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
NORMAL-PM
NORMAL-MATCHED
SMALL 7/8 (88%)
LARGE 2/4 (50%)
SQUAMOUS 10/16 (63%)
ADENOCARCINOMA 4/15 (26%)
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75
Expression relative to median of normal PM samples

FIG. 60

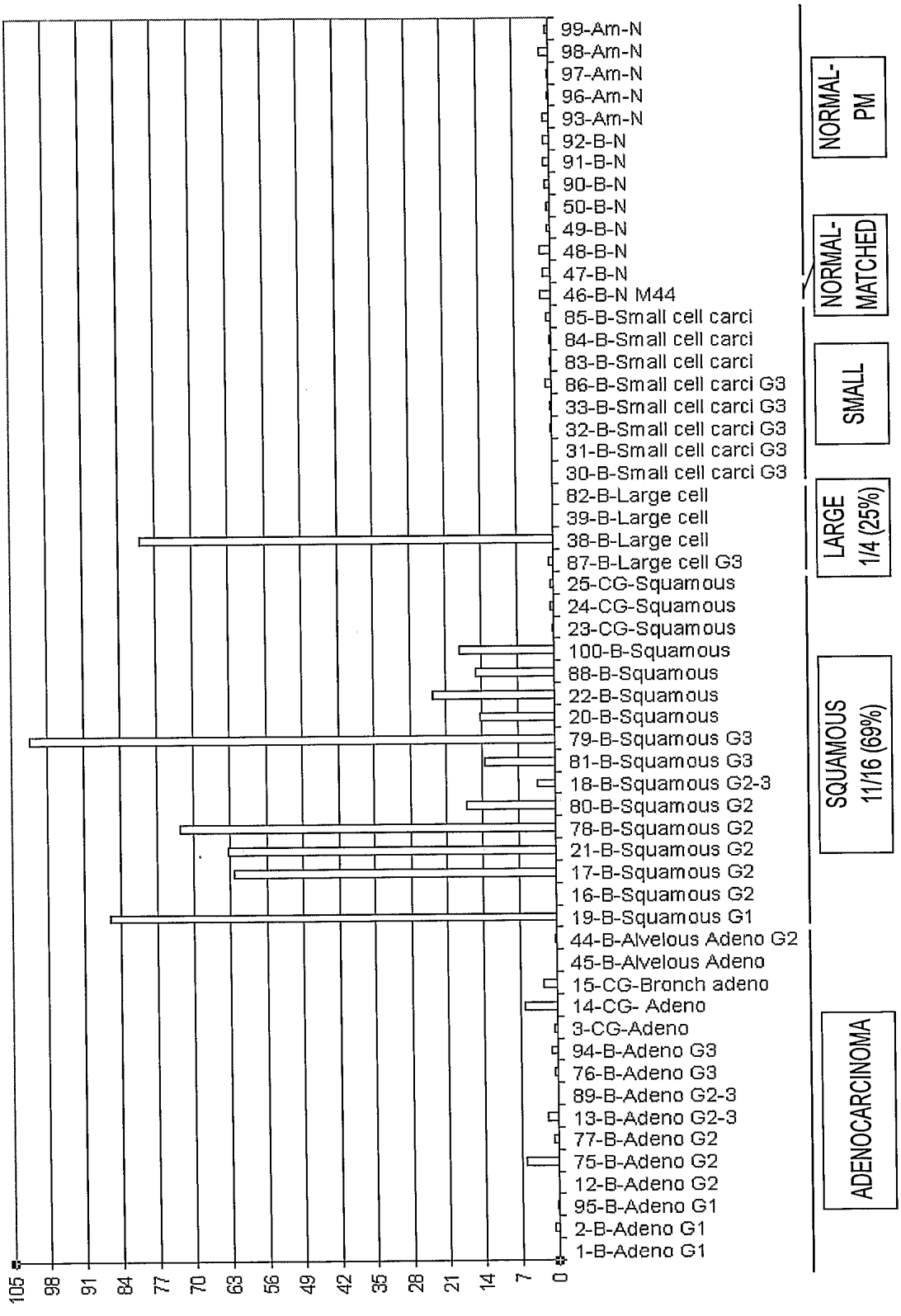

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
105
98
91
84
77
70
63
56
49
42
35
28
21
14
7
0
NORMAL-PM
NORMAL-MATCHED
SMALL
LARGE 1/4 (25%)
SQUAMOUS 11/16 (69%)
ADENOCARCINOMA

FIG. 61

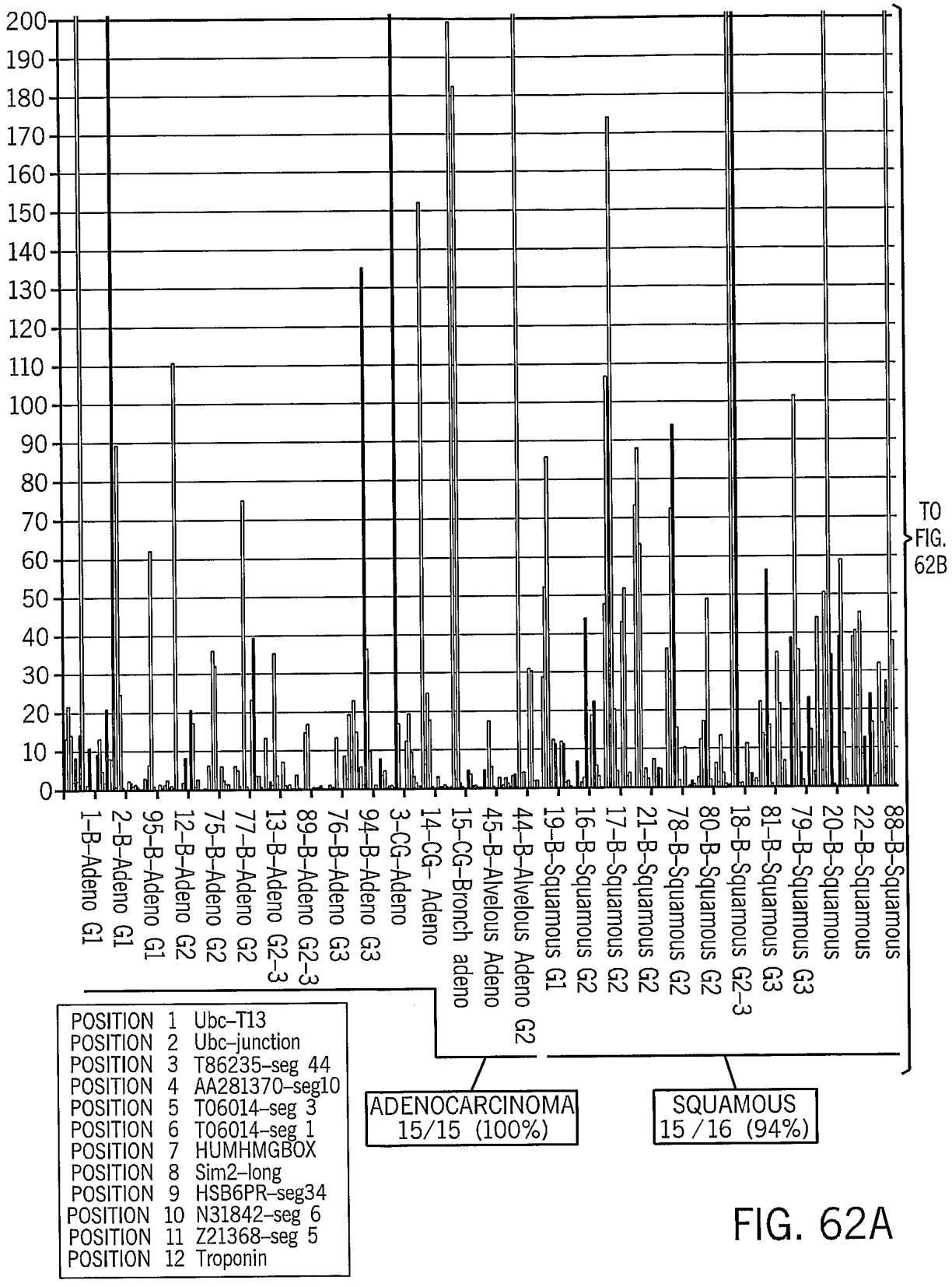

200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
TO FIG. 62B
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alvelous Adeno
44-B-Alvelous Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
POSITION 1 Ubc-T13
POSITION 2 Ubc-junction
POSITION 3 T86235-seg 44
POSITION 4 AA281370-seg10
POSITION 5 T06014-seg 3
POSITION 6 T06014-seg 1
POSITION 7 HUMHMGBOX
POSITION 8 Sim2-long
POSITION 9 HSB6PR-seg34
POSITION 10 N31842-seg 6
POSITION 11 Z21368-seg 5
POSITION 12 Troponin
ADENOCARCINOMA 15/15 (100%)
SQUAMOUS 15 / 16 (94%)

FIG. 62A

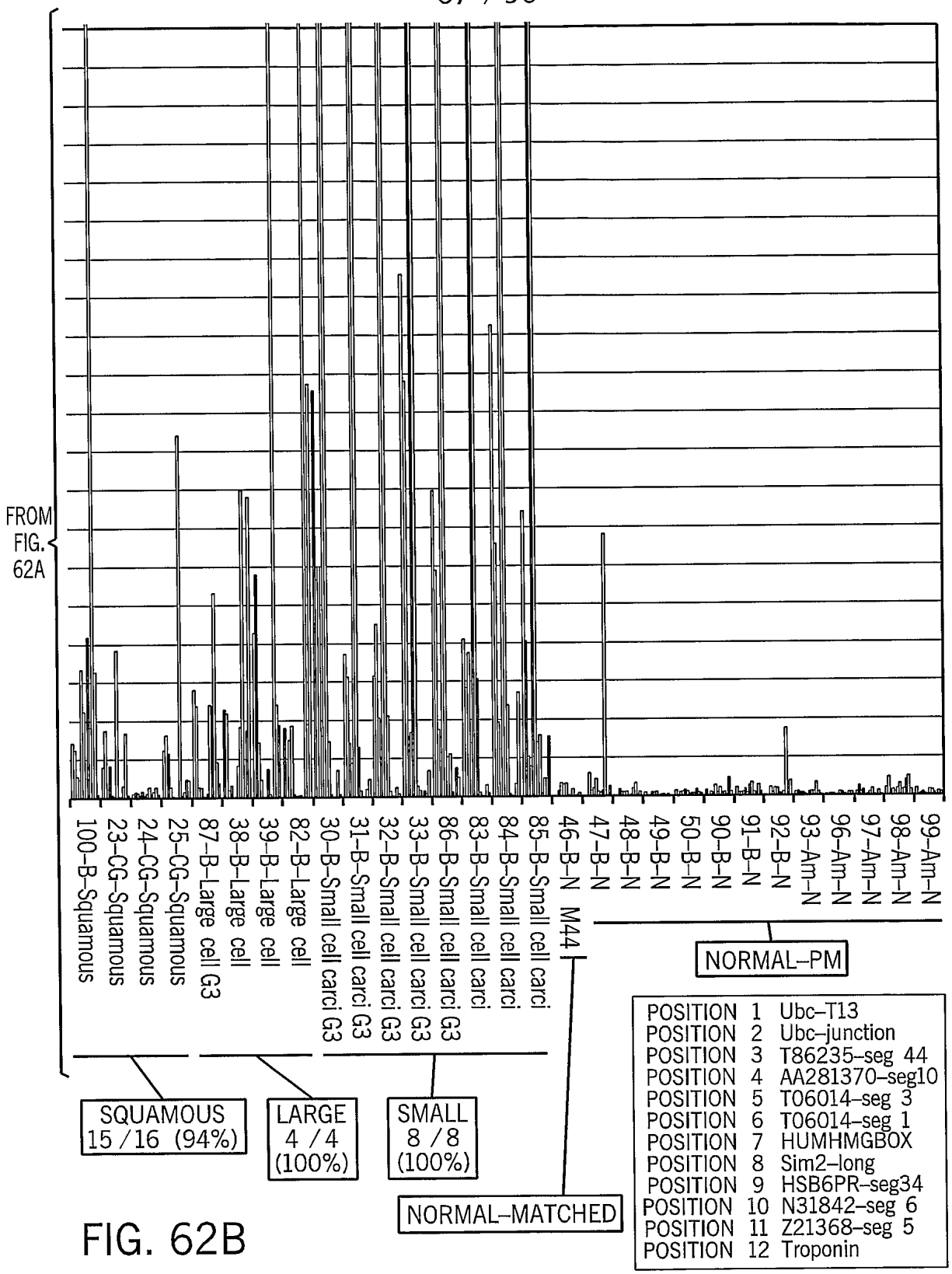

FROM FIG. 62A
100–B–Squamous
23–CG–Squamous
24–CG–Squamous
25–CG–Squamous
87–B–Large cell G3
38–B–Large cell
39–B–Large cell
82–B–Large cell
30–B–Small cell carci G3
31–B–Small cell carci G3
32–B–Small cell carci G3
33–B–Small cell carci G3
86–B–Small cell carci G3
83–B–Small cell carci
84–B–Small cell carci
85–B–Small cell carci
46–B–N M44
47–B–N
48–B–N
49–B–N
50–B–N
90–B–N
91–B–N
92–B–N
93–Am–N
96–Am–N
97–Am–N
98–Am–N
99–Am–N
NORMAL–PM
SQUAMOUS 15 /16 (94%)
LARGE 4 /4 (100%)
SMALL 8 /8 (100%)
NORMAL–MATCHED
POSITION 1 Ubc–T13
POSITION 2 Ubc–junction
POSITION 3 T86235–seg 44
POSITION 4 AA281370–seg10
POSITION 5 T06014–seg 3
POSITION 6 T06014–seg 1
POSITION 7 HUMHMGBOX
POSITION 8 Sim2–long
POSITION 9 HSB6PR–seg34
POSITION 10 N31842–seg 6
POSITION 11 Z21368–seg 5
POSITION 12 Troponin

FIG. 62B

>gi|47124622|gb|AAH70449.1| ☐ Mapkbp1 protein [Mus musculus]

Length = 1503

```
 Score =  911 bits (2354), Expect = 0.0
 Identities = 447/759 (58%), Positives = 576/759 (75%), Gaps = 11/759 (1%)

Query: 40  APPICLRRRTRLSTASEETVQNRVSLEKVLGITAQNSSGLTCDPGTGHVAYLAGCVVVIL 99
           +P I LRR    +    E + ++V+LEKVLG+T      GL CDP +G VAY AGCVVV+
Sbjct: 19  SPSIKLRRSK--AGNRREDLSSKVTLEKVLGVTVSGGRGLACDPRSGLVAYSAGCVVVLF 76

Query: 100 DPKENKQQHIFNTARKSLSALAFSPDGKYIVTGENGHRPAVRIWDVEEKNQVAEMLGHKY 159
           +P+++KQ HI N++RK+++ALAFSPDGKY+VTGE+GH PAVR+WDV E++QVAE+  HKY
Sbjct: 77  NPRKHKQHHILNSSRKTITALAFSPDGKYLVTGESGHMPAVRVWDVAERSQVAELQEHKY 136

Query: 160 GVACVAFSPNMKHIVSMGYQHDMVLNVWDWKKDIVVASNKVSCRVIALSFSEDSSYFVTV 219
           GVACVAFSP+ K+IVS+GYQHDM++NVW WKK+IVVASNKVS RV A+SFSED SYFVT
Sbjct: 137 GVACVAFSPSAKYIVSVGYQHDMIVNVWAWKKNIVVASNKVSSRVTAVSFSEDCSYFVTA 196

Query: 220 GNRHVRFWFLXXXXXXXXXXXXPLVGRSGILGELHNNIFCGVACGRGRMAGSTFCVSYSG 279
           GNRH++FW+L            PL+GRSG+LGEL NN+F  VACGRG  A STFC++ SG
Sbjct: 197 GNRHIKFWYLDDSKTSKVNATVPLLGRSGLLGELRNNLFTDVACGRGEKADSTFCITSSG 256

Query: 280 LLCQFNEKRVLEKWINLKXXXXXXXXXXQELIFCGCTDGIVRIFQAHSLHYLANLPKPHY 339
           LLC+F+++R+L+KW+ L+           QE IFCGC DG VR+F   +LH+L+ LP+PH
Sbjct: 257 LLCEFSDRRLLDKWVELRTTVAHCISVTQEYIFCGCADGTVRLFNPSNLHFLSTLPRPHA 316

Query: 340 LGVDVAQGLEPSFLFHRKAEAVYPDTVALTFDPIHQWLSCVYKDHSIYIWDVKDINRVGK 399
           LG D+A   E S LF     A YPDT+ALTFDP +QWLSCVY DHSIY+WDV+D  +VGK
Sbjct: 317 LGTDIASITEASRLFSGGVNARYPDTIALTFDPTNQWLSCVYNDHSIYVWDVRDPKKVGK 376

Query: 400 VWSELFHSSYVWNVEVYPEFED-QRACLPSGSFLTCSSDNTIRFWNLDSSP--DSHWQKN 456
           V+S L+HSS VW+VEVYPE +D  +ACLP  SF+TCSSDNTIR WN +SS    S   +N
Sbjct: 377 VYSALYHSSCVWSVEVYPEIKDSHQACLPPSSFITCSSDNTIRLWNTESSGVHGSTLHRN 436

Query: 457 IFSNTLLKVVYVENDIQHLQDMSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRS 516
           I SN L+K++YV+ + Q L D +  P     +G+ MD + G+R + +SP+GQHLASGDR
Sbjct: 437 ILSNDLIKIIYVDGNTQALLD-TELPGGDKADGSLMDPRVGIRSVCISPNGQHLASGDRM 495

Query: 517 GNLRIHELHFMDELVKVEAHDAEVLCLEYSKPETGLTLLASASRDRLIHVLNVEKNYNLE 576
           G LRIHEL  + E++KVEAHD+E+LCLEYSKP+TGL LLASASRDRLIHVL+  + Y+L+
Sbjct: 496 GTLRIHELQSLSEMLKVEAHDSEILCLEYSKPDTGLKLLASASRDRLIHVLDAGREYSLQ 555

Query: 577 QTLDDHSSSITAIKFAGNR-DIQMISCGADKSIYFRSAQQGSDGLHFVRTHHVAEKTTLY 635
           QTLD+HSSSITA+KFA +   ++MISCGADKSIYFR+AQ+  +G+ F RTHHV  KTTLY
Sbjct: 556 QTLDEHSSSITAVKFAASDGQVRMISCGADKSIYFRTAQKSGEGVQFTRTHHVVRKTTLY 615

Query: 636 DMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLATSC 695
           DMD++ + KY A+ CQDRN+R++N  +GKQKK +KGSQG++G+L+KV  DPSG ++ATSC
Sbjct: 616 DMDVEPSWKYTAIGCQDRNIRIFNISSGKQKKLFKGSQGEDGTLIKVQTDPSGIYIATSC 675

Query: 696 SDKSISVIDFYSGECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCM 755
           SDK++S+ DF SGEC+A MFGHSEI+T MKF+ DC HLI+VSGDSC+F+W L  E+T  M
Sbjct: 676 SDKNLSIFDFSSGECVATMFGHSEIVTGMKFSNDCKHLISVSGDSCIFVWRLSSEMTISM 735

Query: 756 KQHLLEIDHRQ----QQQHTNDKKRSGHPRQDTYVSTPS 790
           +Q L E+  RQ    QQ  T+ ++ SG  +    V  PS
Sbjct: 736 RQRLAELRQRQRGIKQQGPTSPQRASGAKQHHAPVVPPS 774
```

FIG. 63A

>gi|34856717|ref|XP_342499.1| similar to JNK-binding protein JNKBP1 [Rattus norvegicus]

Length = 1530

Score = 910 bits (2353), Expect = 0.0

Identities = 467/868 (53%), Positives = 611/868 (70%), Gaps = 29/868 (3%)

```
Query: 40  APPICLRRRTRLSTASEETVQNRVSLEKVLGITAQNSSGLTCDPGTGHVAYLAGCVVVIL 99
           +P I LRR    +    E + ++V+LEKVLG+T     GL CDP +G VAY AGCVVV+
Sbjct: 18  SPSIKLRRSK--AGNRREDLSSKVTLEKVLGVTVSGGRGLACDPRSGLVAYPAGCVVVLF 75

Query: 100 DPKENKQQHIFNTARKSLSALAFSPDGKYIVTGENGHRPAVRIWDVEEKNQVAEMLGHKY 159
           +P+++KQ HI N++RK+++ALAFSPDGKY+VTGE+GH PAVR+WDV E+NQVAE+  HKY
Sbjct: 76  NPRKHKQHHILNSSRKTITALAFSPDGKYLVTGESGHMPAVRVWDVAERNQVAELQEHKY 135

Query: 160 GVACVAFSPNMKHIVSMGYQHDMVLNVWDWKKDIVVASNKVSCRVIALSFSEDSSYFVTV 219
           GVACVAFSP+ K+IVS+GYQHDM++NVW WKK+IVVASNKVS RV A+SFSED SYFVT
Sbjct: 136 GVACVAFSPSAKYIVSVGYQHDMIVNVWAWKKNIVVASNKVSSRVTAVSFSEDCSYFVTA 195

Query: 220 GNRHVRFWFLXXXXXXXXXXXXPLVGRSGILGELHNNIFCGVACGRGRMAGSTFCVSYSG 279
           GNRH++FW+L            PL+GRSG+LGEL NN+F  VACGRG+ A STFC++ SG
Sbjct: 196 GNRHIKFWYLDDSKTSKVNATVPLLGRSGLLGELRNNLFTDVACGRGKKADSTFCITSSG 255

Query: 280 LLCQFNEKRVLEKWINLK------XXXXXXXXXXQELIFCGCTDGIVRIFQAHSLHYLAN 333
           LLC+F+++R+L+KW+ L+                QE IFCGC DG VR+F   +LH+L+
Sbjct: 256 LLCEFSDRRLLDKWVELRNTDSFTTTVAHCISVSQEYIFCGCADGTVRLFNPSNLHFLST 315

Query: 334 LPKPHYLGVDVAQGLEPSFLFHRKAEAVYPDTVALTFDPIHQWLSCVYKDHSIYIWDVKD 393
           LP+PH LG D+A   E S LF   A A YPDT+ALTFDP +QWLSCVY DHSIY+WDV+D
Sbjct: 316 LPRPHALGTDIATITEASRLFSGGANARYPDTIALTFDPANQWLSCVYNDHSIYVWDVRD 375

Query: 394 INRVGKVWSELFHSSYVWNVEVYPEFED-QRACLPSGSFLTCSSDNTIRFWNLDSSP--D 450
             +VGKV+S L+HSS VW+VEVYPE +D  +ACLP  SF+TCSSDNTIR WN +SS
Sbjct: 376 PKKVGKVYSALYHSSCVWSVEVYPEIKDSNQACLPPSSFITCSSDNTIRLWNTESSGVHG 435

Query: 451 SHWQKNIFSNTLLKVVYVENDIQHLQDMSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHL 510
           S   +NI SN L+K++YV+ + Q L D +  P     +G+ MD + G+R + +SP+GQHL
Sbjct: 436 SALHRNILSNDLIKIIYVDGNTQALLD-TELPGGDKADGSLMDPRVGIRSVCISPNGQHL 494

Query: 511 ASGDRSGNLRIHELHFMDELVKVEAHDAEVLCLEYSKPETGLTLLASASRDRLIHVLNVE 570
           ASGDR G LR+HEL  + EL+KVEAHD+E+LCLEYSKP+TGL LLASASRDRLIHVL+
Sbjct: 495 ASGDRMGTLRVHELQSLSELLKVEAHDSEILCLEYSKPDTGLKLLASASRDRLIHVLDAG 554

Query: 571 KNYNLEQTLDDHSSSITAIKFAGNR-DIQMISCGADKSIYFRSAQQGSDGLHFVRTHHVA 629
           + Y+L+QTLD+HSSSITA+KFA +   ++MISCGADKSIYFR+AQ+  +G+ F RTHHV
Sbjct: 555 REYSLQQTLDEHSSSITAVKFAASDGQVRMISCGADKSIYFRTAQKSGEGVQFTRTHHVV 614

Query: 630 EKTTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGT 689
            KTTLYDMD++ + KY A+ CQDRN+R++N  +GKQKK +KGSQG++G+L+KV  DPSG
Sbjct: 615 RKTTLYDMDVEPSWKYTAIGCQDRNIRIFNISSGKQKKLFKGSQGEDGTLIKVQTDPSGI 674

Query: 690 FLATSCSDKSISVIDFYSGECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGP 749
           ++ATSCSDK++S+ DF+SGEC+A MFGHSEI+T MKF+ DC HLI+VSGDSC+F+W L
Sbjct: 675 YIATSCSDKNLSIFDFFSGECVATMFGHSEIVTGMKFSNDCKHLISVSGDSCIFVWRLSS 734

Query: 750 EITNCMKQHLLEIDHRQ----QQQHTNDKKRSGHPRQDTYVSTPSEIHSLSPGXXXXXXX 805
           E+T  M+Q L E+  RQ    QQ  T+ +K SG  +    V  PS      P
Sbjct: 735 EMTISMRQRLAELRQRQRGIKQQGPTSPQKASGAKQHHPPVVPPS-----GPALSSDSDK 789

Query: 806 XXXXXXXXMLKTPSKDSLDPDPRCLLTNGKLPL-------WAKRLLGDDDVADGSAFHAK 858
                + P+   L    + L +G  P        W      ++   G A A
Sbjct: 790 EGEDEGTEEEELPALPILGKSTKKELASGSSPALLRSLSHWEMSRAQENMEFLGPAPTAN 849

Query: 859 RSYQPHGRWAERAGQEPLKTILDAQDLD 886
              +  GRWA+   +  ++++LD + L+
Sbjct: 850 TGPKRRGRWAQPGVELSVRSMLDLRQLE 877
```

FIG. 63B

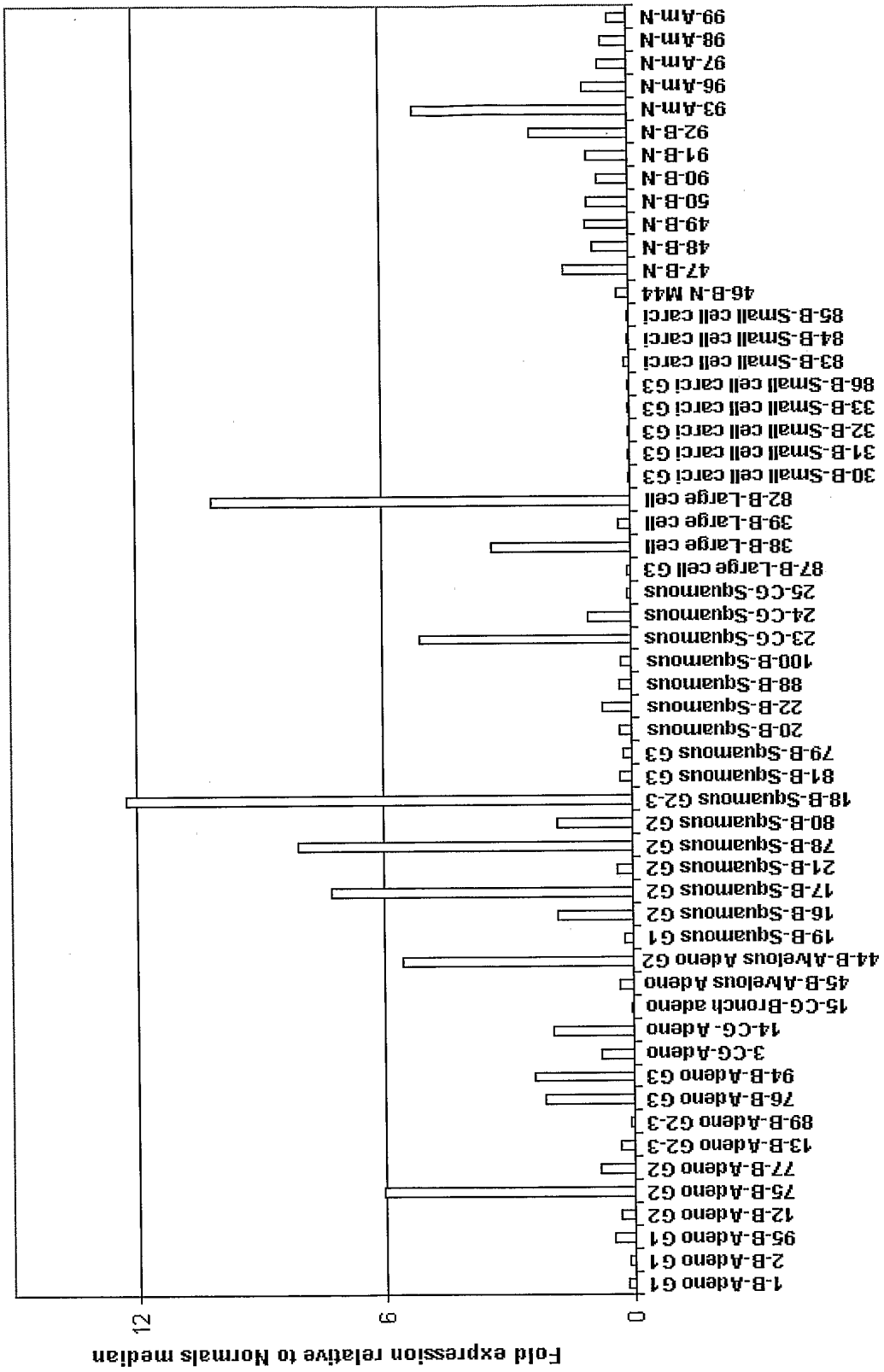

Fold expression relative to Normals median
0
6
12
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Brunch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 64

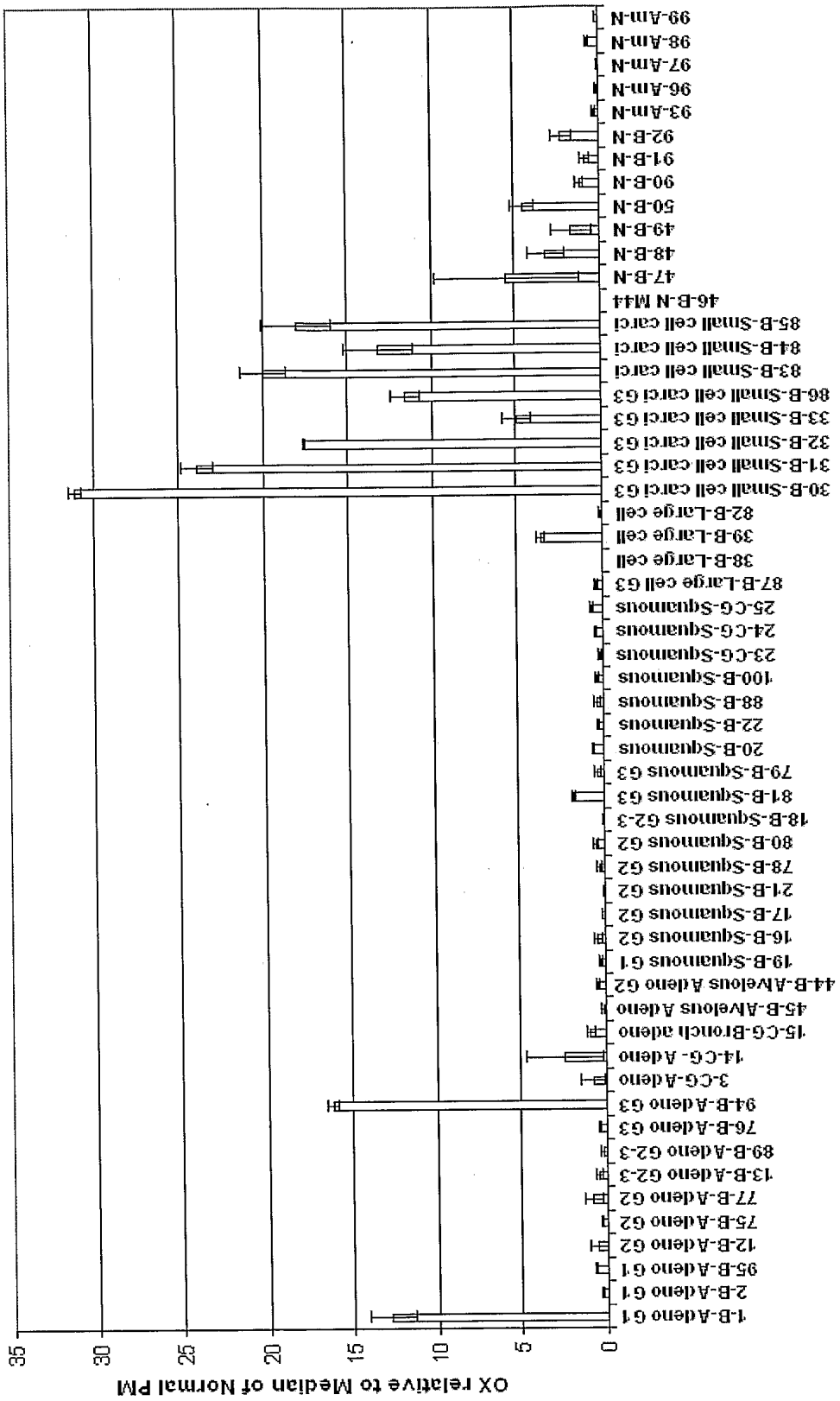

OX relative to Median of Normal PM
0
5
10
15
20
25
30
35
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 65

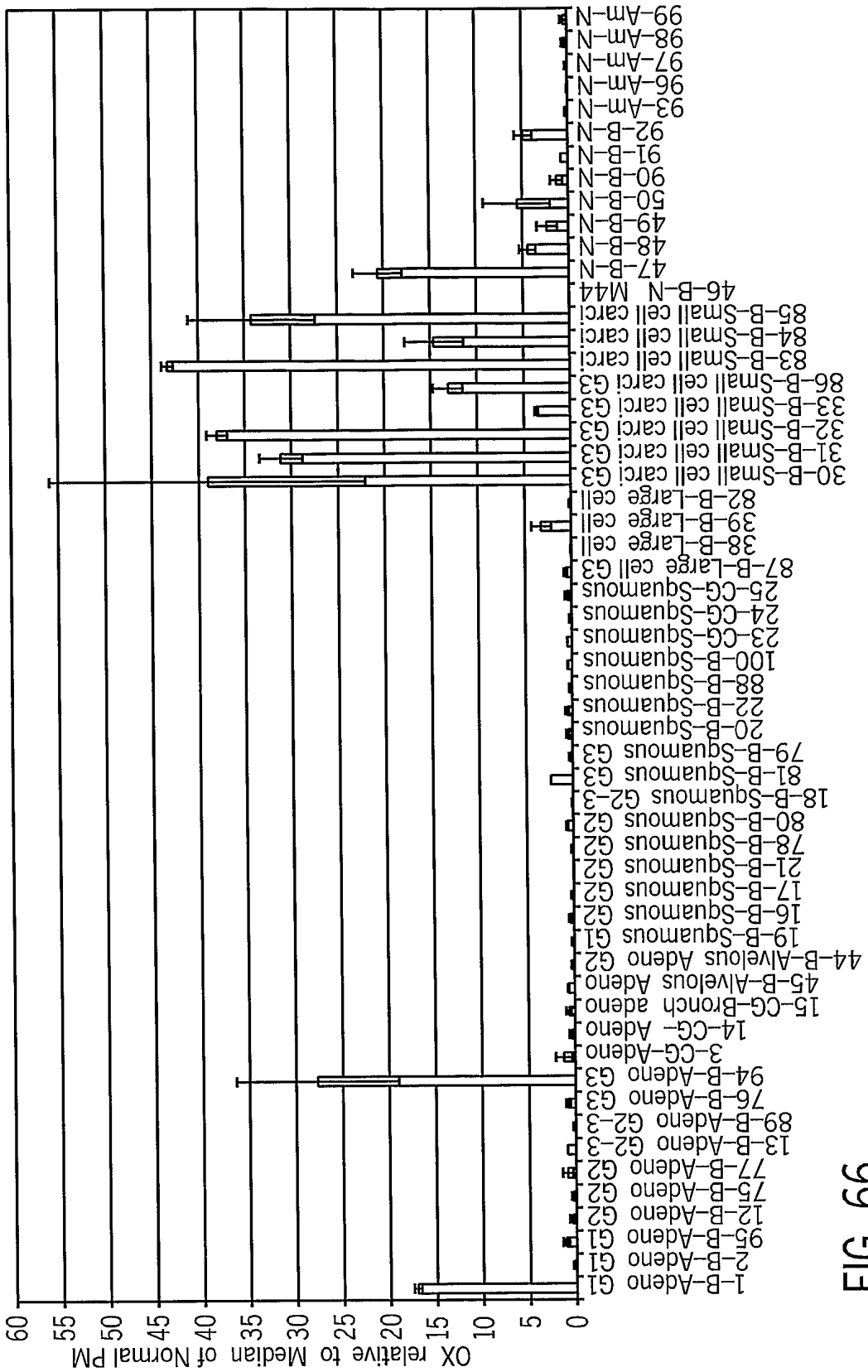

OX relative to Median of Normal PM
0
5
10
15
20
25
30
35
40
45
50
55
60
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 66

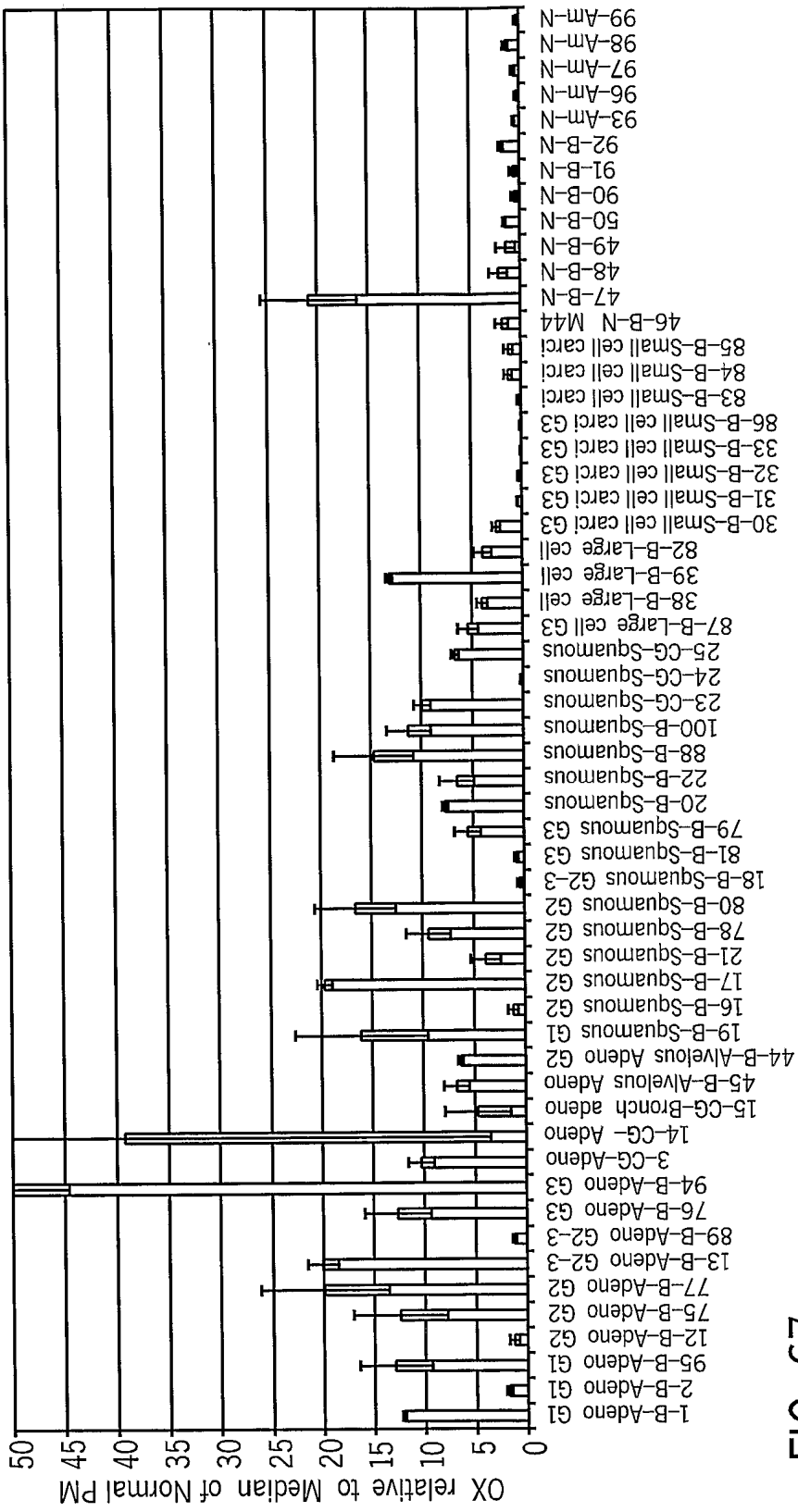

OX relative to Median of Normal PM
0
5
10
15
20
25
30
35
40
45
50
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 67

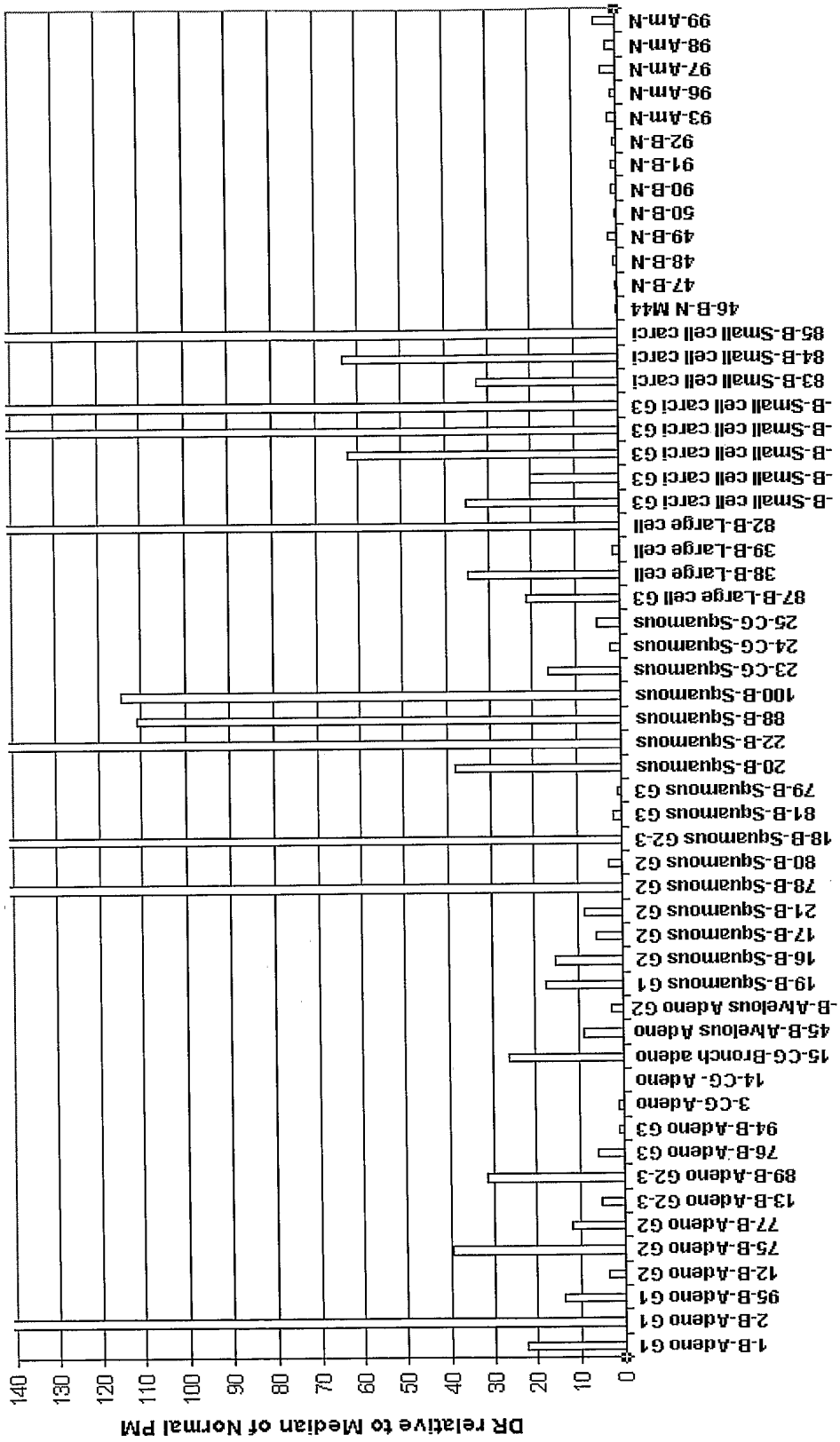

DR relative to Median of Normal PM
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
-B-Small cell carci G3
-B-Small cell carci G3
-B-Small cell carci G3
-B-Small cell carci G3
-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 68

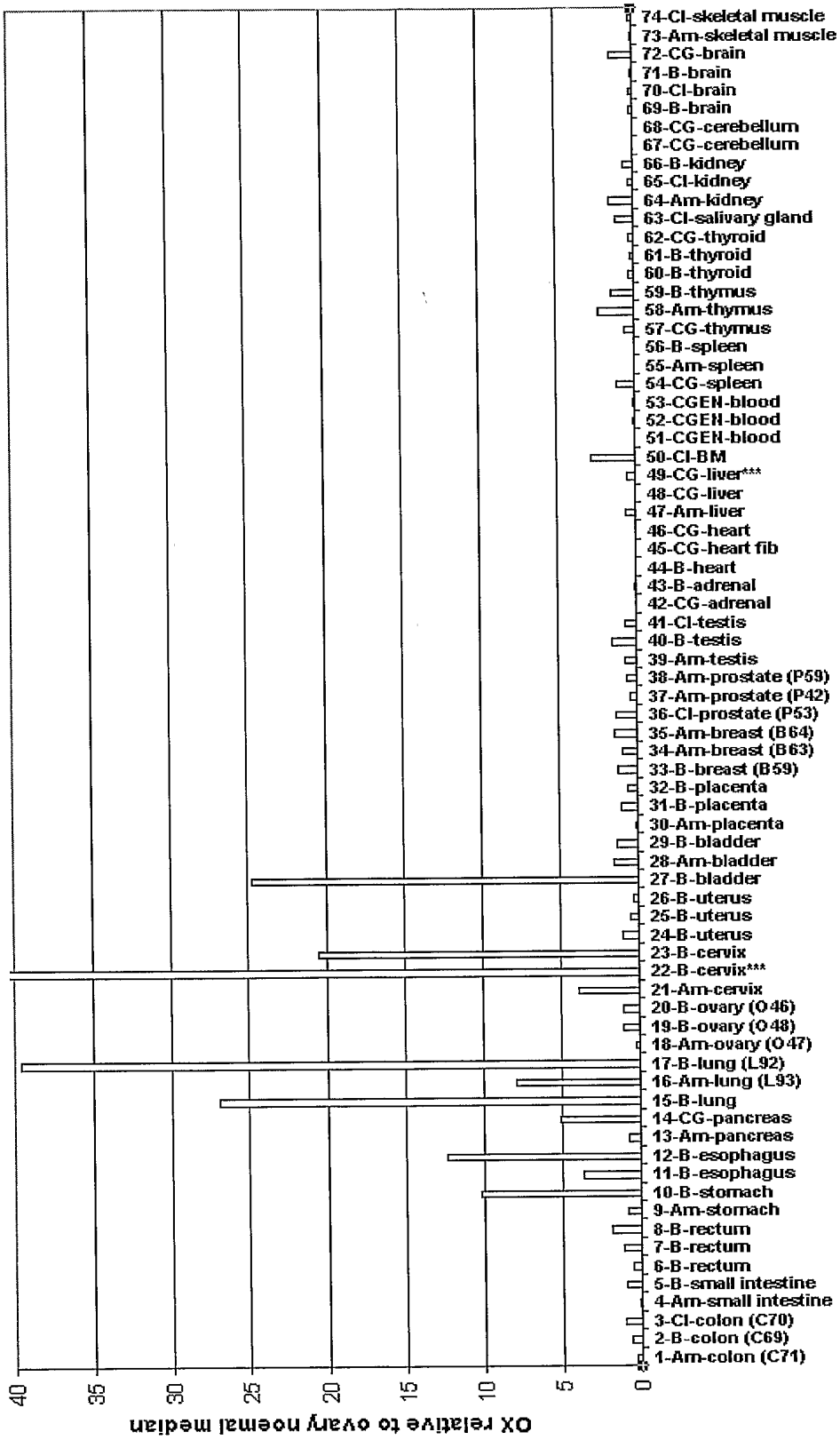

74-Cl-skeletal muscle
73-Am-skeletal muscle
72-CG-brain
71-B-brain
70-Cl-brain
69-B-brain
68-CG-cerebellum
67-CG-cerebellum
66-B-kidney
65-Cl-kidney
64-Am-kidney
63-Cl-salivary gland
62-CG-thyroid
61-B-thyroid
60-B-thyroid
59-B-thymus
58-Am-thymus
57-CG-thymus
56-B-spleen
55-Am-spleen
54-CG-spleen
53-CGEN-blood
52-CGEN-blood
51-CGEN-blood
50-Cl-BM
49-CG-liver***
48-CG-liver
47-Am-liver
46-CG-heart
45-CG-heart fib
44-B-heart
43-B-adrenal
42-CG-adrenal
41-Cl-testis
40-B-testis
39-Am-testis
38-Am-prostate (P59)
37-Am-prostate (P42)
36-Cl-prostate (P53)
35-Am-breast (B64)
34-Am-breast (B63)
33-B-breast (B59)
32-B-placenta
31-B-placenta
30-Am-placenta
29-B-bladder
28-Am-bladder
27-B-bladder
26-B-uterus
25-B-uterus
24-B-uterus
23-B-cervix
22-B-cervix***
21-Am-cervix
20-B-ovary (O46)
19-B-ovary (O48)
18-Am-ovary (O47)
17-B-lung (L92)
16-Am-lung (L93)
15-B-lung
14-CG-pancreas
13-Am-pancreas
12-B-esophagus
11-B-esophagus
10-B-stomach
9-Am-stomach
8-B-rectum
7-B-rectum
6-B-rectum
5-B-small intestine
4-Am-small intestine
3-Cl-colon (C70)
2-B-colon (C69)
1-Am-colon (C71)
0
5
10
15
20
25
30
35
40
OX relative to ovary noemal median

FIG. 69

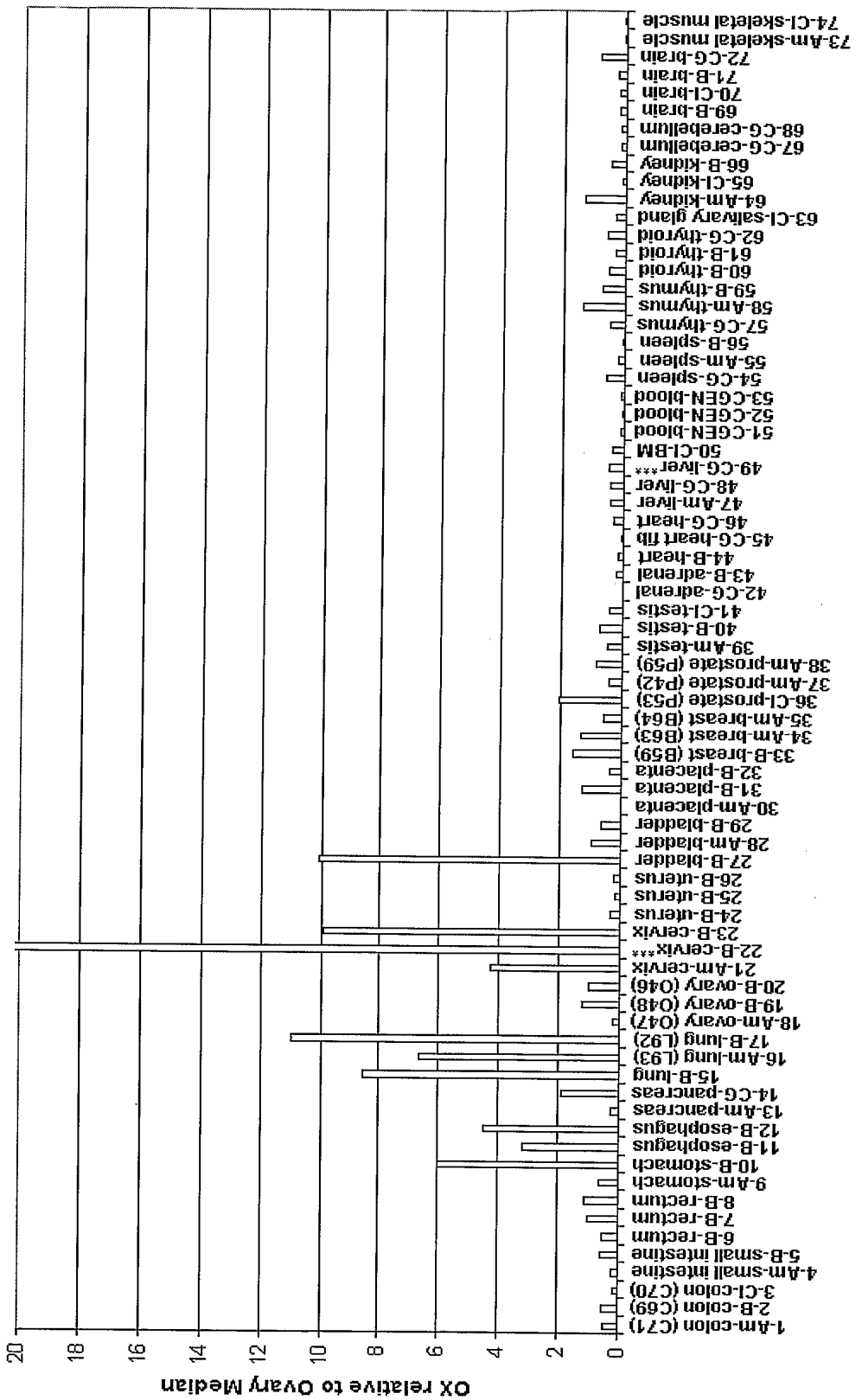

OX relative to Ovary Median
0
2
4
6
8
10
12
14
16
18
20
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart f/b
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 70

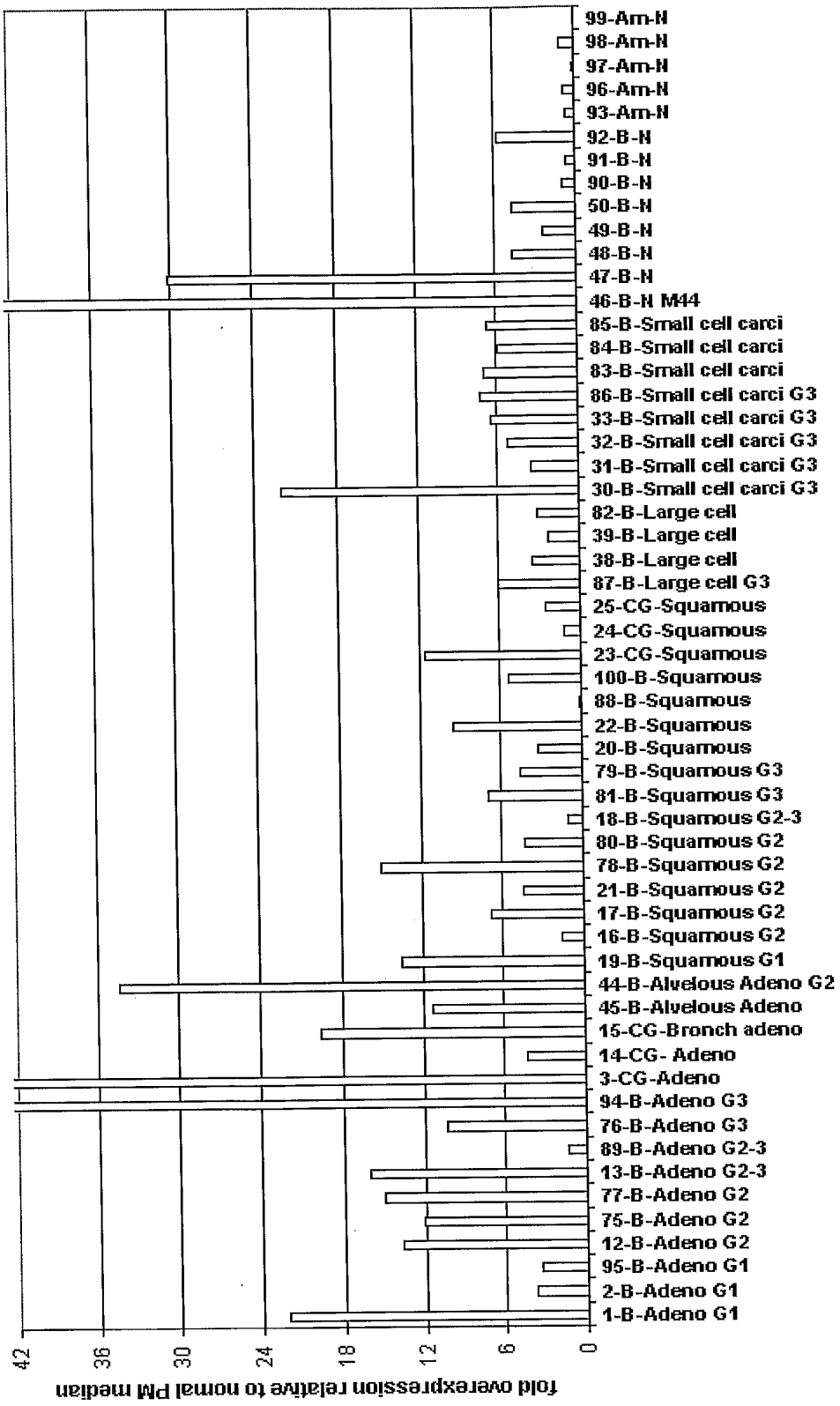

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
42
36
30
24
18
12
6
0
fold overexpression relative to nomal PM median

FIG. 71

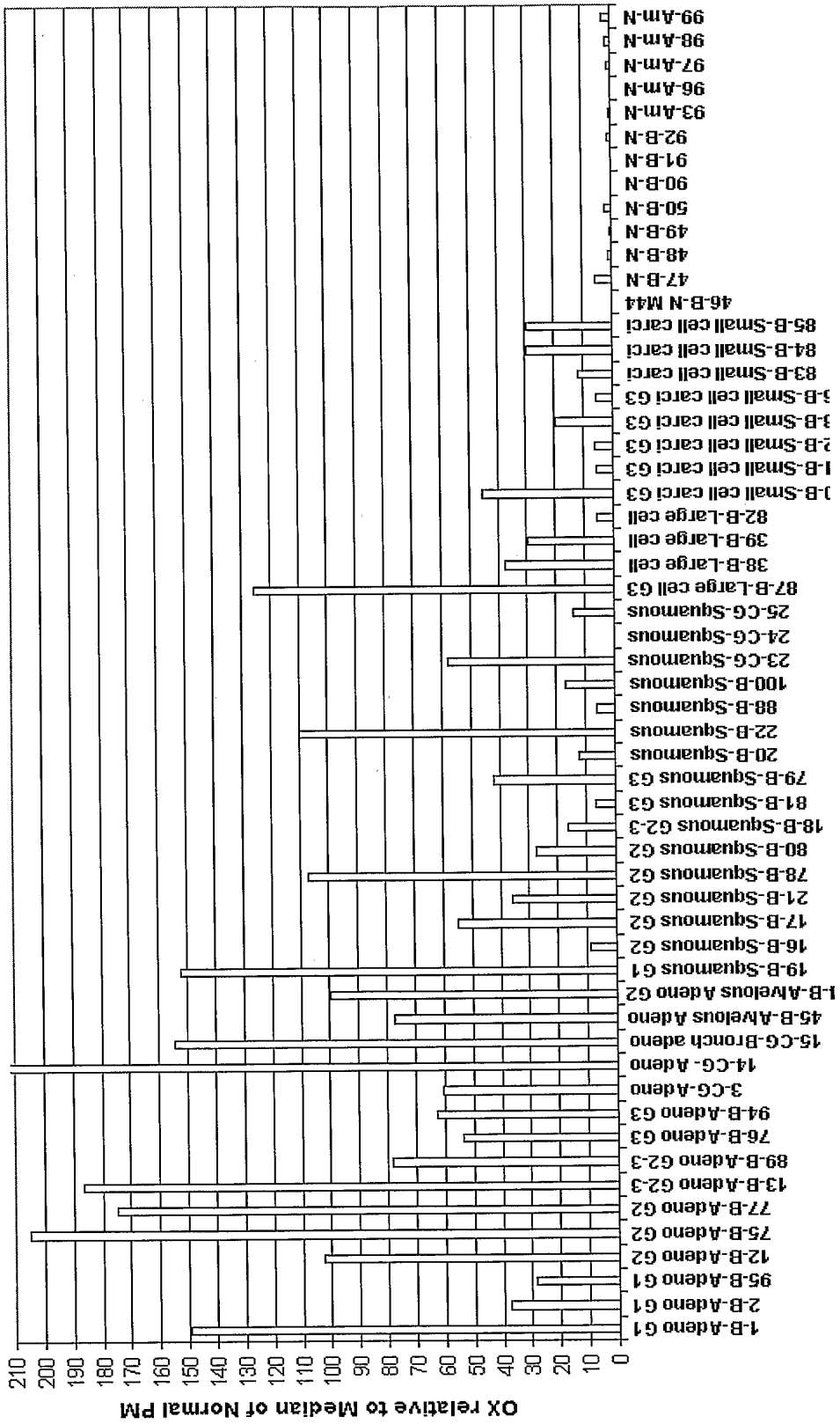

OX relative to Median of Normal PM
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Brunch adeno
45-B-Alveolus Adeno
1-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
)-B-Small cell carci G3
1-B-Small cell carci G3
2-B-Small cell carci G3
3-B-Small cell carci G3
5-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 72

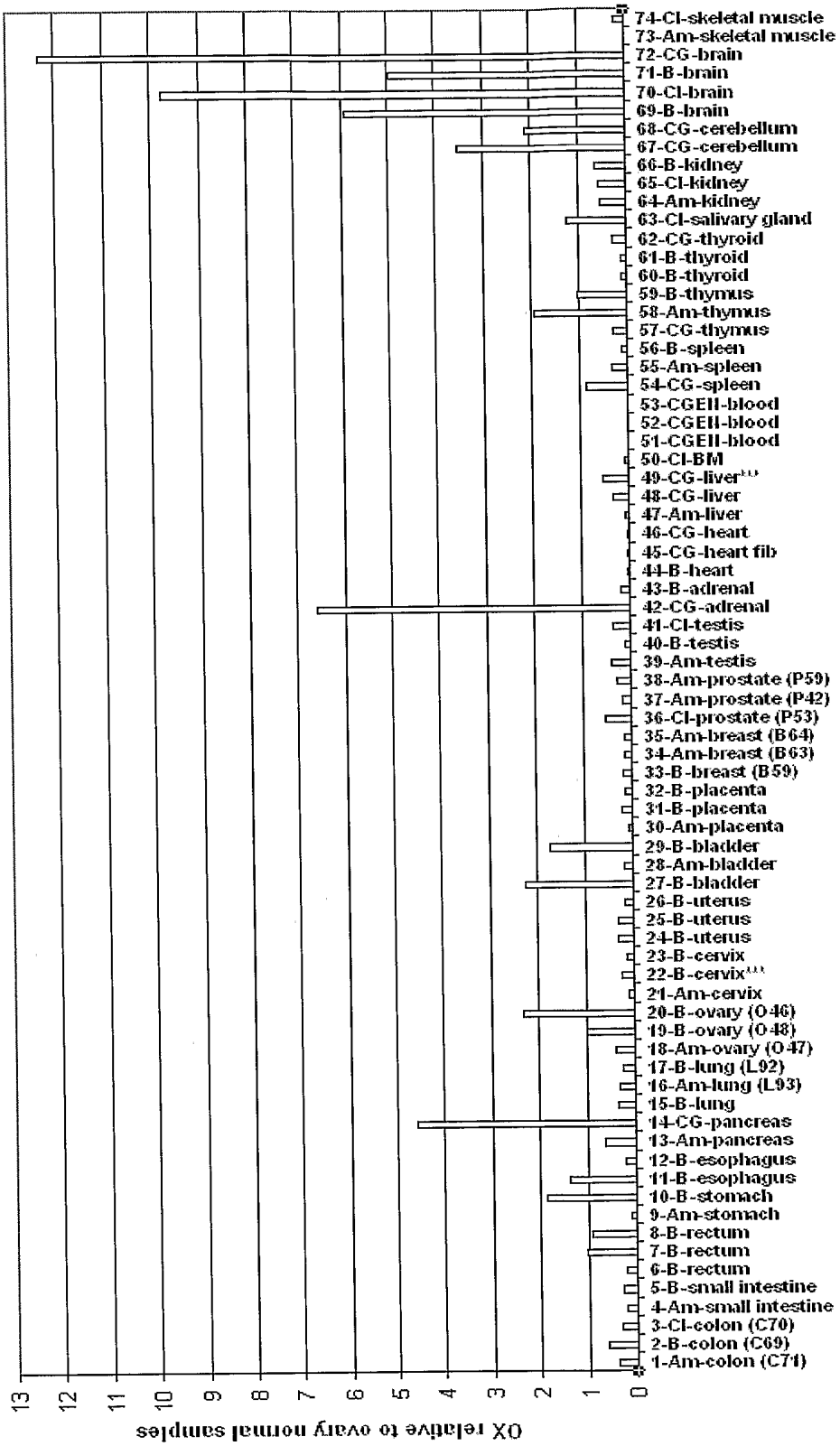

74-Cl-skeletal muscle
73-Am-skeletal muscle
72-CG-brain
71-B-brain
70-Cl-brain
69-B-brain
68-CG-cerebellum
67-CG-cerebellum
66-B-kidney
65-Cl-kidney
64-Am-kidney
63-Cl-salivary gland
62-CG-thyroid
61-B-thyroid
60-B-thyroid
59-B-thymus
58-Am-thymus
57-CG-thymus
56-B-spleen
55-Am-spleen
54-CG-spleen
53-CGEll-blood
52-CGEll-blood
51-CGEll-blood
50-Cl-BM
49-CG-liver(1)
48-CG-liver
47-Am-liver
46-CG-heart
45-CG-heart fib
44-B-heart
43-B-adrenal
42-CG-adrenal
41-Cl-testis
40-B-testis
39-Am-testis
38-Am-prostate (P59)
37-Am-prostate (P42)
36-Cl-prostate (P53)
35-Am-breast (B64)
34-Am-breast (B63)
33-B-breast (B59)
32-B-placenta
31-B-placenta
30-Am-placenta
29-B-bladder
28-Am-bladder
27-B-bladder
26-B-uterus
25-B-uterus
24-B-uterus
23-B-cervix
22-B-cervix(1)
21-Am-cervix
20-B-ovary (O46)
19-B-ovary (O48)
18-Am-ovary (O47)
17-B-lung (L92)
16-Am-lung (L93)
15-B-lung
14-CG-pancreas
13-Am-pancreas
12-B-esophagus
11-B-esophagus
10-B-stomach
9-Am-stomach
8-B-rectum
7-B-rectum
6-B-rectum
5-B-small intestine
4-Am-small intestine
3-Cl-colon (C70)
2-B-colon (C69)
1-Am-colon (C71)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
OX relative to ovary normal samples

FIG. 73

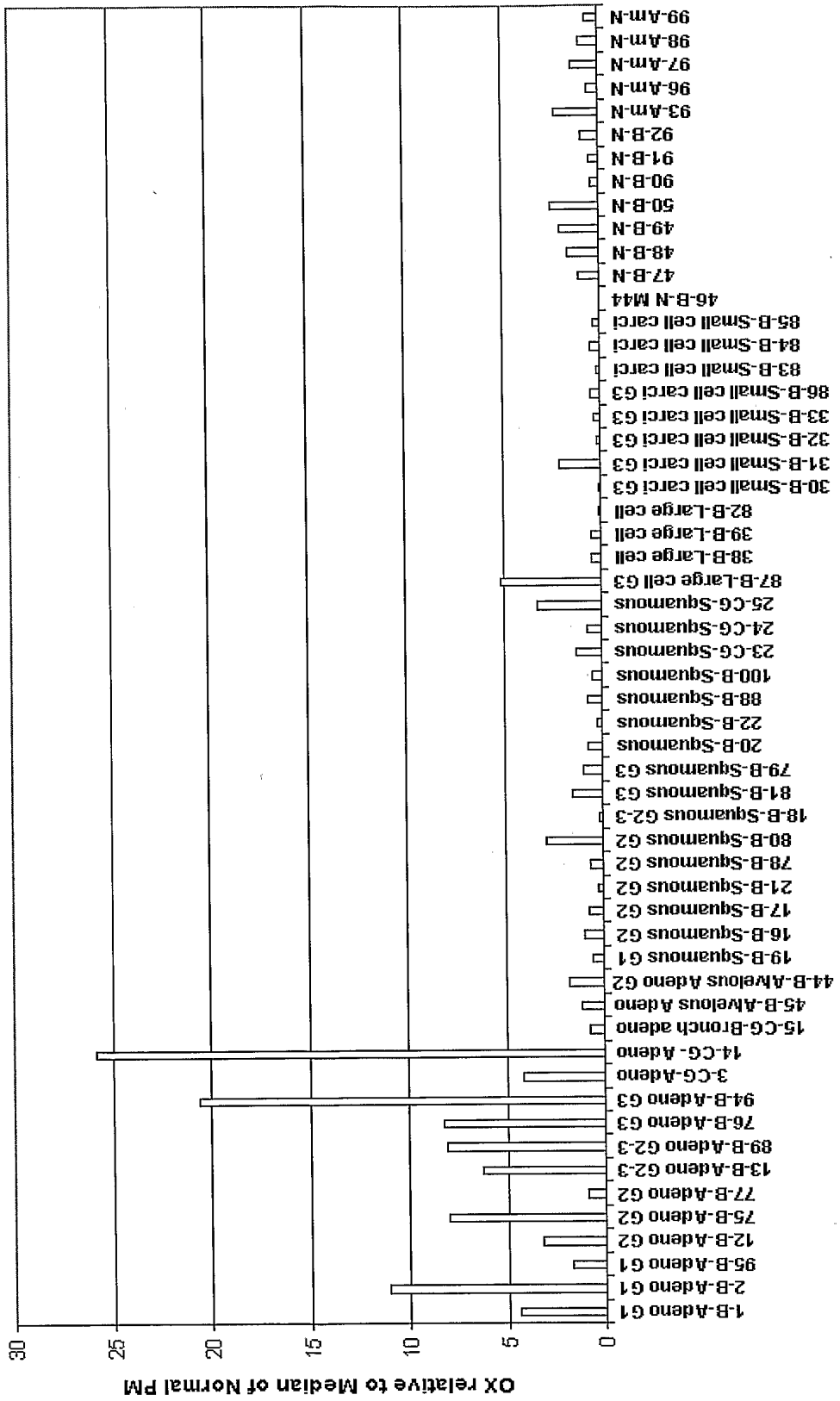

OX relative to Median of Normal PM
0
5
10
15
20
25
30
1-B-Adeno G1
2-B-Adeno G1
95-B-Adeno G1
12-B-Adeno G2
75-B-Adeno G2
77-B-Adeno G2
13-B-Adeno G2-3
89-B-Adeno G2-3
76-B-Adeno G3
94-B-Adeno G3
3-CG-Adeno
14-CG- Adeno
15-CG-Bronch adeno
45-B-Alveolus Adeno
44-B-Alveolus Adeno G2
19-B-Squamous G1
16-B-Squamous G2
17-B-Squamous G2
21-B-Squamous G2
78-B-Squamous G2
80-B-Squamous G2
18-B-Squamous G2-3
81-B-Squamous G3
79-B-Squamous G3
20-B-Squamous
22-B-Squamous
88-B-Squamous
100-B-Squamous
23-CG-Squamous
24-CG-Squamous
25-CG-Squamous
87-B-Large cell G3
38-B-Large cell
39-B-Large cell
82-B-Large cell
30-B-Small cell carci G3
31-B-Small cell carci G3
32-B-Small cell carci G3
33-B-Small cell carci G3
86-B-Small cell carci G3
83-B-Small cell carci
84-B-Small cell carci
85-B-Small cell carci
46-B-N M44
47-B-N
48-B-N
49-B-N
50-B-N
90-B-N
91-B-N
92-B-N
93-Am-N
96-Am-N
97-Am-N
98-Am-N
99-Am-N

FIG. 74

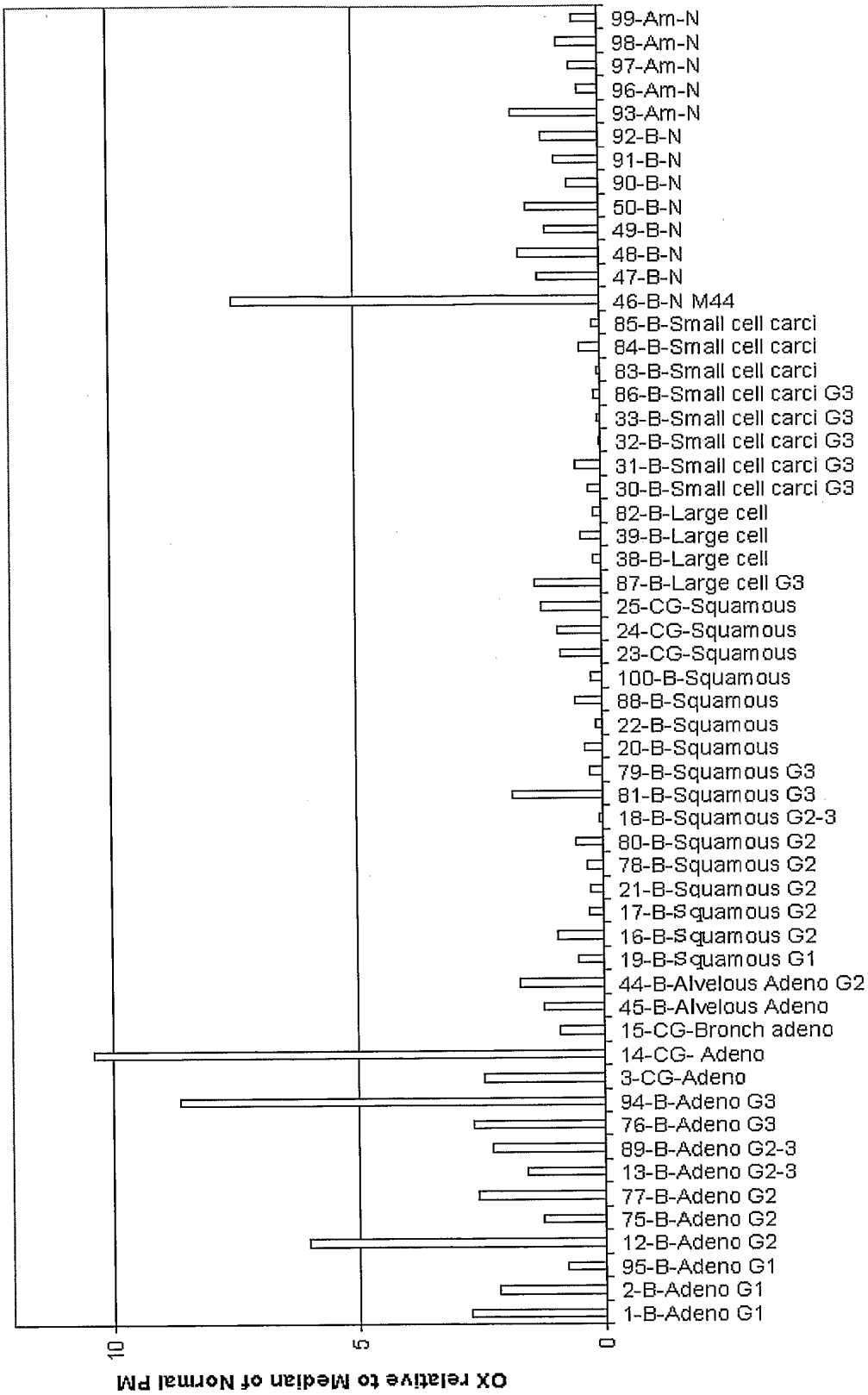

99-Am-N
98-Am-N
97-Am-N
96-Am-N
93-Am-N
92-B-N
91-B-N
90-B-N
50-B-N
49-B-N
48-B-N
47-B-N
46-B-N M44
85-B-Small cell carci
84-B-Small cell carci
83-B-Small cell carci
86-B-Small cell carci G3
33-B-Small cell carci G3
32-B-Small cell carci G3
31-B-Small cell carci G3
30-B-Small cell carci G3
82-B-Large cell
39-B-Large cell
38-B-Large cell
87-B-Large cell G3
25-CG-Squamous
24-CG-Squamous
23-CG-Squamous
100-B-Squamous
88-B-Squamous
22-B-Squamous
20-B-Squamous
79-B-Squamous G3
81-B-Squamous G3
18-B-Squamous G2-3
80-B-Squamous G2
78-B-Squamous G2
21-B-Squamous G2
17-B-Squamous G2
16-B-Squamous G2
19-B-Squamous G1
44-B-Alvelous Adeno G2
45-B-Alvelous Adeno
15-CG-Bronch adeno
14-CG- Adeno
3-CG-Adeno
94-B-Adeno G3
76-B-Adeno G3
89-B-Adeno G2-3
13-B-Adeno G2-3
77-B-Adeno G2
75-B-Adeno G2
12-B-Adeno G2
95-B-Adeno G1
2-B-Adeno G1
1-B-Adeno G1
0
5
10
OX relative to Median of Normal PM

FIG. 75

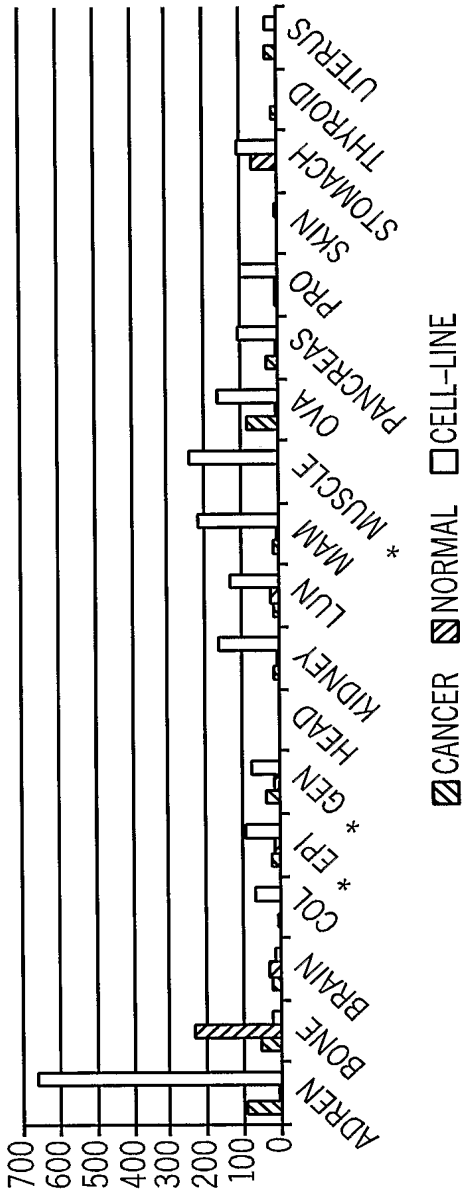

700
600
500
400
300
200
100
0
ADREN
BONE
BRAIN
COL
* EPI
* GEN
HEAD
KIDNEY
LUN
MAM
* MUSCLE
OVA
PANCREAS
PRO
SKIN
STOMACH
THYROID
UTERUS
CANCER
NORMAL
CELL-LINE

FIG. 76

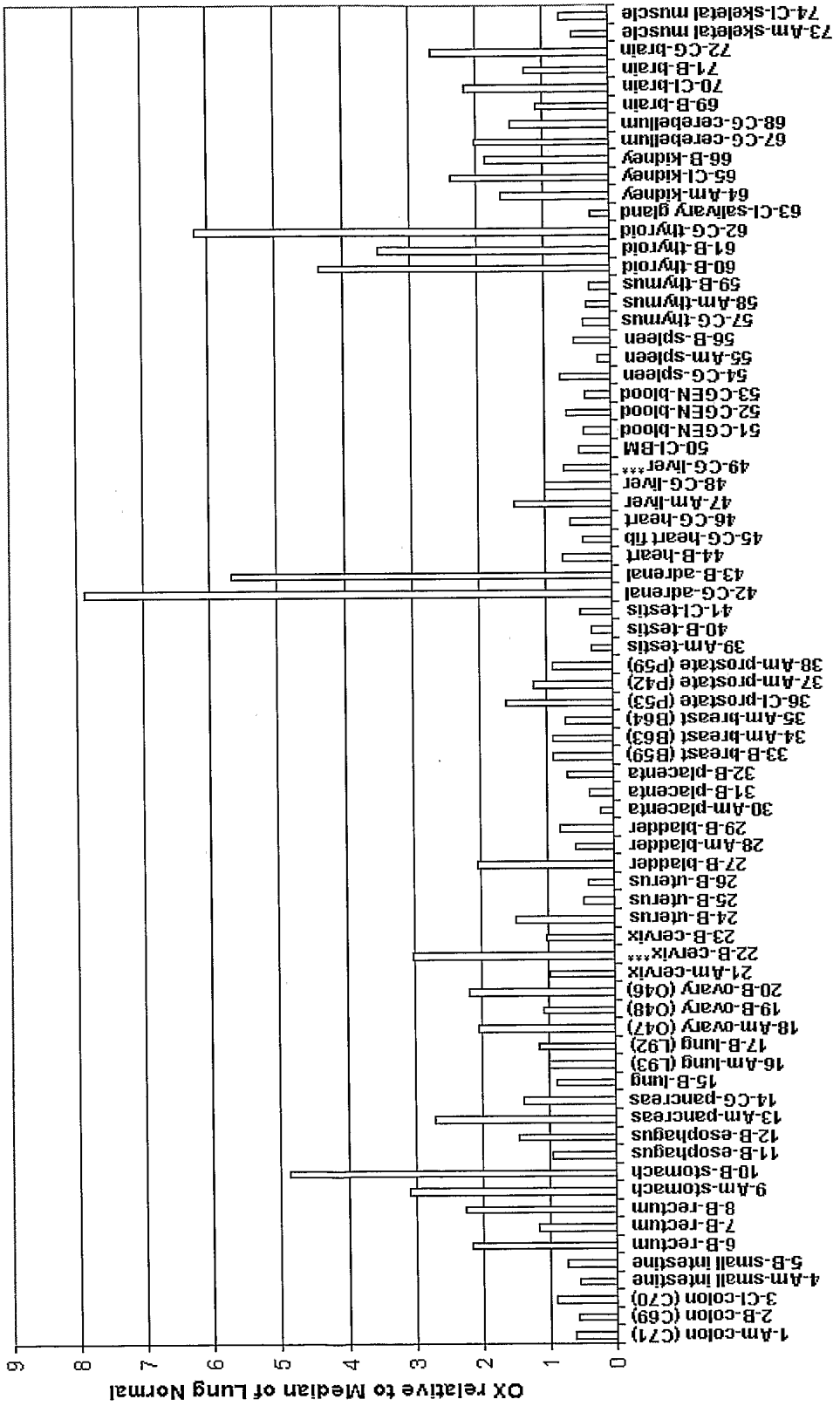

OX relative to Median of Lung Normal
0
1
2
3
4
5
6
7
8
9
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart f/b
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 77

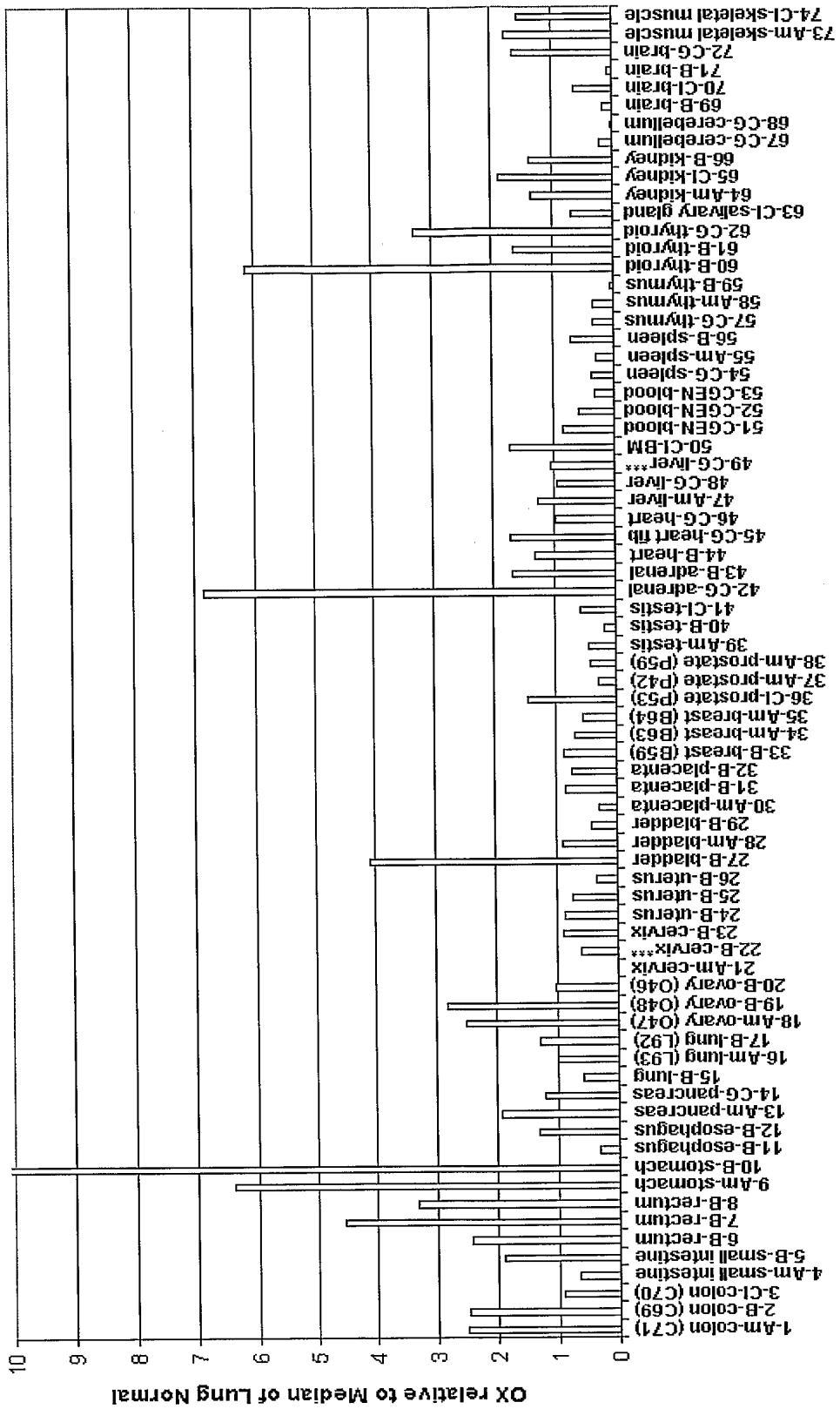

OX relative to Median of Lung Normal
0
1
2
3
4
5
6
7
8
9
10
1-Am-colon (C71)
2-B-colon (C69)
3-Cl-colon (C70)
4-Am-small intestine
5-B-small intestine
6-B-rectum
7-B-rectum
8-B-rectum
9-Am-stomach
10-B-stomach
11-B-esophagus
12-B-esophagus
13-Am-pancreas
14-CG-pancreas
15-B-lung
16-Am-lung (L93)
17-B-lung (L92)
18-Am-ovary (O47)
19-B-ovary (O48)
20-B-ovary (O46)
21-Am-cervix
22-B-cervix***
23-B-cervix
24-B-uterus
25-B-uterus
26-B-uterus
27-B-bladder
28-Am-bladder
29-B-bladder
30-Am-placenta
31-B-placenta
32-B-placenta
33-B-breast (B59)
34-Am-breast (B63)
35-Am-breast (B64)
36-Cl-prostate (P53)
37-Am-prostate (P42)
38-Am-prostate (P59)
39-Am-testis
40-B-testis
41-Cl-testis
42-CG-adrenal
43-B-adrenal
44-B-heart
45-CG-heart f/b
46-CG-heart
47-Am-liver
48-CG-liver
49-CG-liver***
50-Cl-BM
51-CGEN-blood
52-CGEN-blood
53-CGEN-blood
54-CG-spleen
55-Am-spleen
56-B-spleen
57-CG-thymus
58-Am-thymus
59-B-thymus
60-B-thyroid
61-B-thyroid
62-CG-thyroid
63-Cl-salivary gland
64-Am-kidney
65-Cl-kidney
66-B-kidney
67-CG-cerebellum
68-CG-cerebellum
69-B-brain
70-Cl-brain
71-B-brain
72-CG-brain
73-Am-skeletal muscle
74-Cl-skeletal muscle

FIG. 78

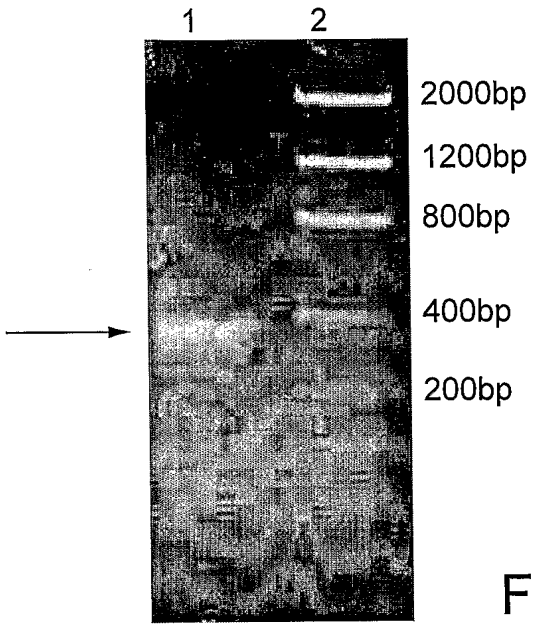


FIG. 79

TGCTGTCGCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGAC
TCCGGGAGA***ATGTGGGTCCTAGGCATCGCGGCAACTTTTTGCGGAT***
TGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGT
GAAGAATTCCAGCTGAACAACGACTGCTCCTCCCCCGAGTTCATTG
TGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGAT
GGAGCAAAGTGCCGACACTAAAAGAACAAACACCTTGCTCTTCGA
GATGAGACATTTTGCCAAGCAGTTGACCACTTAGTTCTCAAGAAGCA
ACTATCT*CTTTCATGTGCCTTCTGAGG*

FIG. 80

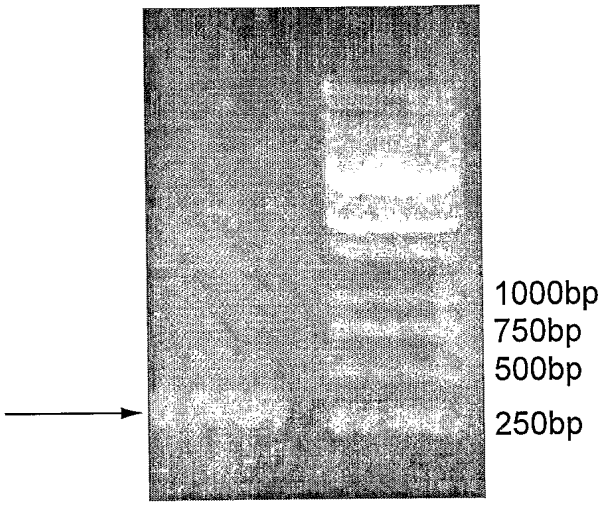


FIG. 81

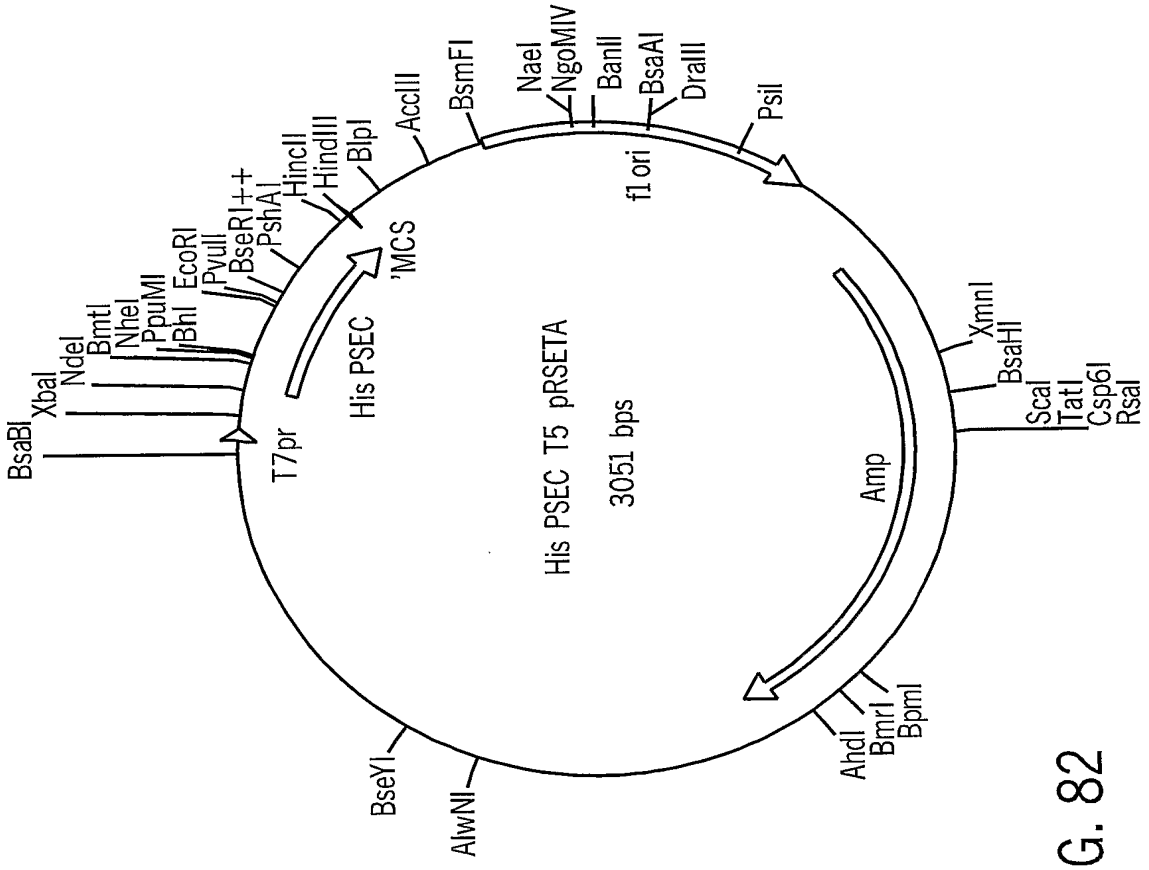

His PSEC T5 pRSETA
3051 bps
T7pr
His PSEC
'MCS
f1 ori
Amp
BsaBI
XbaI
NdeI
BmtI
NheI
PpuMI
BhI
EcoRI
PvuII
BseRI++
PshAI
HincII
HindIII
BlpI
AccIII
BsmFI
NaeI
NgoMIV
BanII
BsaAI
DraIII
PsiI
XmnI
BsaHI
ScaI
TatI
Csp6I
RsaI
AhdI
BmrI
BpmI
AlwNI
BseYI

FIG. 82

MRGS**HHHHHH**GMAS*MWVLGIAATFCGLFLLPGFALQIQCYQCEEF*
QLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSADTKRTNTLLFEMR
HFAKQLTT

FIG. 83

GATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGA*GACCACAACGGTTTCCCTCTAGAA*
*ATAATTTTGTTTAACTTTAAGAAGGAGATATACAT***ATGCGGGGTTCTCATCATCATCATCATCATG**
GTATGGCTAGCATGTGGGTCCTAGGCATCGCGGCAACTTTTTGCGGATTGTTCTTGCTTCCAG
GCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACGACTGCTCC
TCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGAT
GGAGCAAAGTGCCGACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCA
AGCAGTTGACCACTTAG*AAGCTTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTG*
GCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG
GGTTTTTTGCTGAAAGGAGGAACTATATCCGGATCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCA
TTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGC
*GCCCGCTCC*TTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGC
TCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAA
CTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGA
CGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATC
TCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTG
ATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCG
GGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTT
CCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACG
CTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGAT
CTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTT
TTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCG
CCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGG
ATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA
CTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGA
TCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCG
TGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTT
ACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTC
TGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGT
CTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC
GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACT
GATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCAT
TTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTT
TTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTT
GCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACC
AAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTA
CATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTAC
CGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTT
CGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAG
CTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG
GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTC
CTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGA
GCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG
CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGA
GCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGG
AAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAG

FIG. 84

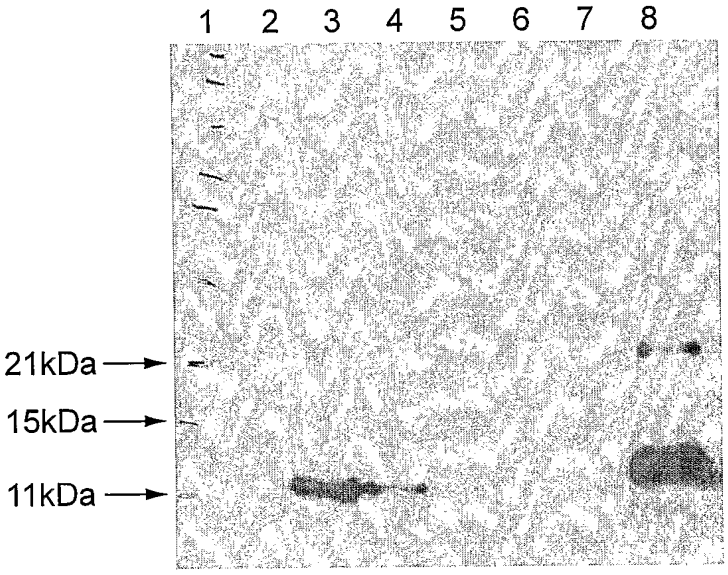


FIG. 85

Sequence Listing:

<210> SEQ ID NO 1

<211> Length : 1,250

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 1

>H61775_T21

GGAGGCGCTCGGGGCATCCGAGGCGGGGAGGCGGGTCCGCCCCCTATTGTGTAGCGGCGAGAGTGGAGCCGAGCGGT
GCGGAGCAGATCTGGTGGTTCTCCGGAGAGCAGCTTCCTTGGGTGTTACATGAGCCAAGCCCTCACTGTACAGAAGA
GTGAGAGCTGAAACCTGTTCCCTGAGCTGATCAGAAGGACATCCCTTGGCCCCTCCATCTGGGCTCCTGTGGATAGG
AGGGGCTGGGTGAGCAGGCCAGCTGGGCTATGGTGTGGTGCCTCGGCCTGGCCGTCCTCAGCCTGGTCATCAGCCAG
GGGGCTGACGGTCGAGGGAAGCCTGAGGTGGTATCGGTGGTGGGCCGGGCTGGGGAGAGTGTGGTGCTGGGCTGTGA
CCTGCTGCCCCCGGCCGGCCGGCCCCCCCTGCATGTCATCGAGTGGCTGCGCTTTGGATTCCTGCTTCCCATCTTCA
TCCAGTTCGGCCTCTACTCTCCCCGAATTGACCCTGATTACGTGGGTGACTGTGGGTTCCCTGCCTTCCGAGAGCTT
AAGAGAGCAGAGACTGTGTCTCCTGTTTTCTTCACACGCCGCTGCATATGGGAAGATCTGAAGTCAACAGGCTTTAG
CCCTGCAGGTGGAGGGAGGCCTCCAGGAGGTGGGCCCAGGACTCAGGAGGACTCAGGGCTGCCCTGCTGGCGATCTT
CCTGTTCTGTAACACTACAGGTCTAGCAGTCCAGCTGTCACAGAAAAGCTAGGACATGCAGTATGCTTCTTTGGATA
TTCTGAGTAACATTTGGACTGTTACCCATTGGCTACCAGCATCTCCCAAGTGAGAATACATAGATTACCCCCAGTGC
CCTGAACAGCACTCGGTCCTAACACCCGTGTCCATGGAAAGCACGCCGCGTCTGGAGAAAGAAGCCGAAGGCTCTTG
TCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATTCCTTCAGCTACAATGGAATTCTTGGGAGGATT
AAATATGGTGACAACGCCTAATATTAGATGGCCTGTATTCCACACTCAATCTTCCTTCCCTCTTCTTCCTTCTTTGT
AGAGCTATAATGAAAAGTATCATGTGGGACACAGAAGAGGTTGCAGTCTGGGGTCTGCAGGGCTTAGCGGCCAGGCA
GATTAGCTTTCTTGAGGAATCCTGACAGTGGGTGGAAGGGTATGATGATG

<210> SEQ ID NO 2

<211> Length : 715

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 2

>H61775_T22

GGAGGCGCTCGGGGCATCCGAGGCGGGGAGGCGGGTCCGCCCCCTATTGTGTAGCGGCGAGAGTGGAGCCGAGCGGT
GCGGAGCAGATCTGGTGGTTCTCCGGAGAGCAGCTTCCTTGGGTGTTACATGAGCCAAGCCCTCACTGTACAGAAGA
GTGAGAGCTGAAACCTGTTCCCTGAGCTGATCAGAAGGACATCCCTTGGCCCCTCCATCTGGGCTCCTGTGGATAGG
AGGGGCTGGGTGAGCAGGCCAGCTGGGCTATGGTGTGGTGCCTCGGCCTGGCCGTCCTCAGCCTGGTCATCAGCCAG
GGGGCTGACGGTCGAGGGAAGCCTGAGGTGGTATCGGTGGTGGGCCGGGCTGGGGAGAGTGTGGTGCTGGGCTGTGA
CCTGCTGCCCCCGGCCGGCCGGCCCCCCCTGCATGTCATCGAGTGGCTGCGCTTTGGATTCCTGCTTCCCATCTTCA
TCCAGTTCGGCCTCTACTCTCCCCGAATTGACCCTGATTACGTGGGATAAGAGTTCTCCTCAGGTGGGAGGTAGGGA
GGTATCAGCAAGAAAGGTGGGCTGGGTAGAGTCGCACAAGGCCTCCTATGAACGGCTTTGTCCCTGCTCTGATCTCA
TCTCCAGCTCTGCTGCCTTAACTCTGCTTAATAAGCATGGCTGTGCTCCCAAGCAGTGTTAATTCATTGAAAGATGT
CATTCATTTACACACACACACA

<210> SEQ ID NO 3

<211> Length : 2,875

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 3

>M85491_PEA_1_T16

TCTGCTGGCTGCGCGGTGGCGGCGGCTGTGTGTGCGCCGCGCCTTGCCGCCCCCCCCTGGCCCCCCGAGCCCGGGGCG
CGCGCTCCCGCCCGGGCCGTCCGGGCCCCGCGGCGCCGCGGCCCGAGGCCCCGGGAAGCGCAGCCATGGCTCTGCGG
AGGCTGGGGGCCGCGCTGCTGCTGCTGCCGCTGCTCGCCGCCGTGGAAGAAACGCTAATGGACTCCACTACAGCGAC
TGCTGAGCTGGGCTGGATGGTGCATCCTCCATCAGGGTGGGAAGAGGTGAGTGGCTACGATGAGAACATGAACACGA
TCCGCACGTACCAGGTGTGCAACGTGTTTGAGTCAAGCCAGAACAACTGGCTACGGACCAAGTTTATCCGGCGCCGT
GGCGCCCACCGCATCCACGTGGAGATGAAGTTTTCGGTGCGTGACTGCAGCAGCATCCCCAGCGTGCCTGGCTCCTG
CAAGGAGACCTTCAACCTCTATTACTATGAGGCTGACTTTGACTCGGCCACCAAGACCTTCCCCAACTGGATGGAGA
ATCCATGGGTGAAGGTGGATACCATTGCAGCCGACGAGAGCTTCTCCCAGGTGGACCTGGGTGGCCGCGTCATGAAA
ATCAACACCGAGGTGCGGAGCTTCGGACCTGTGTCCCGCAGCGGCTTCTACCTGGCCTTCCAGGACTATGGCGGCTG
CATGTCCCTCATCGCCGTGCGTGTCTTCTACCGCAAGTGCCCCCGCATCATCCAGAATGGCGCCATCTTCCAGGAAA
CCCTGTCGGGGGCTGAGAGCACATCGCTGGTGGCTGCCCGGGGCAGCTGCATCGCCAATGCGGAAGAGGTGGATGTA
CCCATCAAGCTCTACTGTAACGGGGACGGCGAGTGGCTGGTGCCCATCGGGCGCTGCATGTGCAAAGCAGGCTTCGA
GGCCGTTGAGAATGGCACCGTCTGCCGAGGTTGTCCATCTGGGACTTTCAAGGCCAACCAAGGGGATGAGGCCTGTA
CCCACTGTCCCATCAACAGCCGGACCACTTCTGAAGGGGCCACCAACTGTGTCTGCCGCAATGGCTACTACAGAGCA

GACCTGGACCCCCTGGACATGCCCTGCACAACCATCCCCTCCGCGCCCCAGGCTGTGATTTCCAGTGTCAATGAGAC
CTCCCTCATGCTGGAGTGGACCCCTCCCCGCGACTCCGGAGGCCGAGAGGACCTCGTCTACAACATCATCTGCAAGA
GCTGTGGCTCGGGCCGGGGTGCCTGCACCCGCTGCGGGGACAATGTACAGTACGCACCACGCCAGCTAGGCCTGACC
GAGCCACGCATTTACATCAGTGACCTGCTGGCCCACACCCAGTACACCTTCGAGATCCAGGCTGTGAACGGCGTTAC
TGACCAGAGCCCCTTCTCGCCTCAGTTCGCCTCTGTGAACATCACCACCAACCAGGCAGCTCCATCGGCAGTGTCCA
TCATGCATCAGGTGAGCCGCACCGTGGACAGCATTACCCTGTCGTGGTCCCAGCCAGACCAGCCCAATGGCGTGATC
CTGGACTATGAGCTGCAGTACTATGAGAAGGTACCTATTGGCTGGGTGCTGTCCCCATCACCCACCTCCCTGAGGGC
CCCTCTCCCAGGCTGAGGCCTGGGAGTTCTGCCCCACCGCAAGATGAGACGCACTGGTGCAGCAGAAAGAGCACTGG
CCTTGGAGTCAGGCTGCCTGGCTCCCAATCCAGCTCCGCTCCTTCCCACTGTGAGACCTCAGGCAGGTGCCTTGACC
TCTCTGGATCTCACTTTTCTGGTCTGGAGGATACACCCAGCAATCTCAGTGAAATGCAACAGTCACATCCCTTTCCC
TACCACGACCCTTTCATCTTGACCTCAGTGGCTTGATGTTGGGAAAAACTGGGTTTCCAAAAAGCTGCACTTATGAA
GTGATAATTAGTCACTCACCTCTTCTTCGACAGAGATTTGAAACAGCTCAAGAGAGCTTCCGCCTGCCCTGCTCTGA
GTCCTGCTAAAACACCCACTTTCACTCGCCTGCATGCCCTTTGCATGGGGAGAGGTGATTTCACTTTGAGCTTTTAA
ATCAGACCTTAATTACTCCCTTTGGGTGGAAGCCCCTGGGATGGTAGAAGGATCACTGGACTAAGAGTGAGAAGCCG
TAGGTTCAAATCCCAGCTCCGTCCTTCACCAGCTATGTGACCTTGGGCAGGCGTCTTTCTCCCTCTGAACCTCAGTT
TCCACCTGTGTCGAGTGTGGGTGAGACCCCTCGCGGGGAGCTATGCAGGTTACGGAGAAAAGGCAGCACAGCACCCA
GAATGGGACCTGGCCCTCAGCAGAGGCCATGTGTGTCCCTGGCCTTCCTCCTCTGCCCTGCCTGCTGCACAGTGGGC
AATGGTGACAGGATGGGAGGCCAAGTGGATGTGGGGTCTGCACAGTACAGGGGCCAGGAGGTAGACAGCACAATTGC
CCACCCACATGGCTGGACATCAGAGGCCCCAGGAAGCCTCTCCTTTGAATGATCACTTCTCTTACCTGCTCCAGGAG
GCAACAAACAGCCACAGAGGCTGCAAGGGCACCTGGGAAAGGCATCGCGGGGCTTCCATTCAGACTAGGTGTCAATG
ACTGACAGGGAGGCCTTTGGTTGAGGGCAAGCCCACGGGGAACTGCAGATGGATGGAAGGGCTCTCCCTGAAGGCTG
AGAGGAAGAGTGCAGTCAATTGCAGCCAGTCCTGCTGGAGCCCAACTTTCTAGAGCCCAGCCCGGCCTTCCCACTCT
GTTAACTGCTGGATCGGCTAACCAGGCCGGTCTCCAGGGCCTTTCAAACACTTACCCAGCCTTTGCCGGCCGTCTTA
CCATTGCTTGCGTGCGTGTTCATCCC

<210> SEQ ID NO 4

<211> Length : 1,182

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 4

>M85491_PEA_1_T20

TCTGCTGGCTGCGCGGTGGCGGCGGCTGTGTGTGCGCCGCGCCTTGCCGCCCCCCCCTGGCCCCCCGAGCCCGGGGCG
CGCGCTCCCGCCCGGGCCGTCCGGGCCCCGCGGCGCCGCGGCCCGAGGCCCCGGGAAGCGCAGCCATGGCTCTGCGG
AGGCTGGGGGCCGCGCTGCTGCTGCTGCCGCTGCTCGCCGCCGTGGAAGAAACGCTAATGGACTCCACTACAGCGAC

TGCTGAGCTGGGCTGGATGGTGCATCCTCCATCAGGGTGGGAAGAGGTGAGTGGCTACGATGAGAACATGAACACGA
TCCGCACGTACCAGGTGTGCAACGTGTTTGAGTCAAGCCAGAACAACTGGCTACGGACCAAGTTTATCCGGCGCCGT
GGCGCCCACCGCATCCACGTGGAGATGAAGTTTTCGGTGCGTGACTGCAGCAGCATCCCCAGCGTGCCTGGCTCCTG
CAAGGAGACCTTCAACCTCTATTACTATGAGGCTGACTTTGACTCGGCCACCAAGACCTTCCCCAACTGGATGGAGA
ATCCATGGGTGAAGGTGGATACCATTGCAGCCGACGAGAGCTTCTCCCAGGTGGACCTGGGTGGCCGCGTCATGAAA
ATCAACACCGAGGTGCGGAGCTTCGGACCTGTGTCCCGCAGCGGCTTCTACCTGGCCTTCCAGGACTATGGCGGCTG
CATGTCCCTCATCGCCGTGCGTGTCTTCTACCGCAAGTGCCCCCGCATCATCCAGAATGGCGCCATCTTCCAGGAAA
CCCTGTCGGGGGCTGAGAGCACATCGCTGGTGGCTGCCCGGGGCAGCTGCATCGCCAATGCGGAAGAGGTGGATGTA
CCCATCAAGCTCTACTGTAACGGGGACGGCGAGTGGCTGGTGCCCATCGGGCGCTGCATGTGCAAAGCAGGCTTCGA
GGCCGTTGAGAATGGCACCGTCTGCCGAGAGAGACAGGATCTCACTATGTTGTCCAGGCTGGTCTTGAACTCGTGGC
CACAAATGATCCTCCCACCTCAGCCTCCCAAAGTGTTGGAATTATAGGCATGAACCACCATGCCCAGGAGGAGAATT
TTTGATAATAATATTTTGTGGACATCTTTGCATATCATGTCAGAGCTATAACATCATTGTGGAGAAGCTCTTAGGAT
CCCATAGAATAAATGTACCGTAATTTA

<210> SEQ ID NO 5

<211> Length : 2,199

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 5

>T39971_T10

GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC
TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGGCTGACCAAGAGTCATGCAAGGGCCGCTGCACTGAGGG
CTTCAACGTGGACAAGAAGTGCCAGTGTGACGAGCTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTG
AGTGCAAGCCCCAAGTGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAG

GAGAAAAACAATGCCACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGG
GAATCCTGAGCAGACACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGG
GGATAGACTCAAGGCCTGAGACCCTTCATCCAGGGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCTCTTTGCCTTCCGAGGGCAGTACTGCTATGAACTGGACGA
AAAGGCAGTGAGGCCTGGGTACCCCAAGCTCATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCA
CCCGCATCAACTGTCAGGGGAAGACCTACCTCTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGTGTCCTGGAC
CCTGATTACCCCCGAAATATCTCTGACGGCTTCGATGGCATCCCGGACAACGTGGATGCAGCCTTGGCCCTCCCTGC
CCATAGCTACAGTGGCCGGGAGCGGGTCTACTTCTTCAAGGGGAAACAGTACTGGGAGTACCAGTTCCAGCACCAGC
CCAGTCAGGAGGAGTGTGAAGGCAGCTCCCTGTCGGCTGTGTTTGAACACTTTGCCATGATGCAGCGGGACAGCTGG
GAGGACATCTTCGAGCTTCTCTTCTGGGGCAGAACCTCTGGCATGGCACCCCGCCCCTCCTTGGCCAAGAAACAAAG
GTTTAGGCATCGCAACCGCAAAGGCTACCGTTCACAACGAGGCCACAGCCGTGGCCGCAACCAGAACTCCCGCCGGC
CATCCCGCGCCACGTGGCTGTCCTTGTTCTCCAGTGAGGAGAGCAACTTGGGAGCCAACAACTATGATGACTACAGG
ATGGACTGGCTTGTGCCTGCCACCTGTGAACCCATCCAGAGTGTCTTCTTCTTCTCTGGAGACAAGTACTACCGAGT
CAATCTTCGCACACGGCGAGTGGACACTGTGGACCCTCCCTACCCACGCTCCATCGCTCAGTACTGGCTGGGCTGCC
CAGCTCCTGGCCATCTGTAGGAGTCAGAGCCCACATGGCCGGGCCCTCTGTAGCTCCCTCCTCCCATCTCCTTCCCC
CAGCCCAATAAAGGTCCCTTAGCCCCGAAAAAAAAGCKATAAT

<210> SEQ ID NO 6

<211> Length : 1,947

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 6

>T39971_T12

GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC
TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGGCTGACCAAGAGTCATGCAAGGGCCGCTGCACTGAGGG
CTTCAACGTGGACAAGAAGTGCCAGTGTGACGAGCTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTG

AGTGCAAGCCCCAAGTGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAG
GAGAAAAACAATGCCACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGG
GAATCCTGAGCAGACACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGG
GGATAGACTCAAGGCCTGAGACCCTTCATCCAGGGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCTCTTTGCCTTCCGAGGGCAGTACTGCTATGAACTGGACGA
AAAGGCAGTGAGGCCTGGGTACCCCAAGCTCATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCA
CCCGCATCAACTGTCAGGGGAAGACCTACCTCTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGTGTCCTGGAC
CCTGATTACCCCCGAAATATCTCTGACGGCTTCGATGGCATCCCGGACAACGTGGATGCAGCCTTGGCCCTCCCTGC
CCATAGCTACAGTGGCCGGGAGCGGGTCTACTTCTTCAAGGGGAAACAGTACTGGGAGTACCAGTTCCAGCACCAGC
CCAGTCAGGAGGAGTGTGAAGGCAGCTCCCTGTCGGCTGTGTTTGAACACTTTGCCATGATGCAGCGGGACAGCTGG
GAGGACATCTTCGAGCTTCTCTTCTGGGGCAGAACCTCTGACAAGTACTACCGAGTCAATCTTCGCACACGGCGAGT
GGACACTGTGGACCCTCCCTACCCACGCTCCATCGCTCAGTACTGGCTGGGCTGCCCAGCTCCTGGCCATCTGTAGG
AGTCAGAGCCCACATGGCCGGGCCCTCTGTAGCTCCCTCCTCCCATCTCCTTCCCCCAGCCCAATAAAGGTCCCTTA
GCCCCGAAAAAAAAGCKATAAT

<210> SEQ ID NO 7

<211> Length : 1,592

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 7

>T39971_T16

GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC
TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGGCTGACCAAGAGTCATGCAAGGGCCGCTGCACTGAGGG
CTTCAACGTGGACAAGAAGTGCCAGTGTGACGAGCTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTG
AGTGCAAGCCCCAAGTGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAG
GAGAAAAACAATGCCACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGG

GAATCCTGAGCAGACACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGG
GGATAGACTCAAGGCCTGAGACCCTTCATCCAGGGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCTCTTTGCCTTCCGAGGGCAGTACTGCTATGAACTGGACGA
AAAGGCAGTGAGGCCTGGGTACCCCAAGCTCATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCA
CCCGCATCAACTGTCAGGGGAAGACCTACCTCTTCAAGGTGCCAGGGGCTGTGGGCCAGGGTAGAAAGCATCTAGGG
AGGGTTTGAGAGCTATTGCTCCCAGGGACAGGGTGGACAGGGAAGCTGGACCCAGGGCCCTGCAGGACCTGGTGGGA
GCTCTGTGAGCACAGGGCAGCCCCAAGACTCCAGGTCCTGGGCAGTGAACCT

<210> SEQ ID NO 8

<211> Length : 2,490

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 8

>T39971_T5

GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC
TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGGCTGACCAAGAGTCATGCAAGGGCCGCTGCACTGAGGG
CTTCAACGTGGACAAGAAGTGCCAGTGTGACGAGCTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTG
AGTGCAAGCCCCAAGTGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAG
GAGAAAAACAATGCCACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGG
GAATCCTGAGCAGACACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGG
GGATAGACTCAAGGCCTGAGACCCTTCATCCAGGGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCTCTTTGCCTTCCGAGGGCAGTACTGCTATGAACTGGACGA
AAAGGCAGTGAGGCCTGGGTACCCCAAGCTCATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCA
CCCGCATCAACTGTCAGGGGAAGACCTACCTCTTCAAGGGTAGTCAGTACTGGCGCTTTGAGGATGGTGTCCTGGAC
CCTGATTACCCCCGAAATATCTCTGACGGCTTCGATGGCATCCCGGACAACGTGGATGCAGCCTTGGCCCTCCCTGC
CCATAGCTACAGTGGCCGGGAGCGGGTCTACTTCTTCAAGGGTACTCAGGGGGTGGTGGGAGACTGAGCAGGCAGTG

GAGCAGTCTTGGATTCCTTTCACATTTCACTGGGGACAGGCCTCAGCATGTGCCCACCCCTGACCCCCACCTCATGC
TGGGAGATCCTAACTTCAACAGCCTCTGGGATCTCCAGTCTTGCCCTGGCCCAGCCCTCCTAATGCCCACCACCCCG
CTCCTCAGGGAAACAGTACTGGGAGTACCAGTTCCAGCACCAGCCCAGTCAGGAGGAGTGTGAAGGCAGCTCCCTGT
CGGCTGTGTTTGAACACTTTGCCATGATGCAGCGGGACAGCTGGGAGGACATCTTCGAGCTTCTCTTCTGGGGCAGA
ACCTCTGCTGGTACCAGACAGCCCCAGTTCATTAGCCGGGACTGGCACGGTGTGCCAGGGCAAGTGGACGCAGCCAT
GGCTGGCCGCATCTACATCTCAGGCATGGCACCCCGCCCCTCCTTGGCCAAGAAACAAAGGTTTAGGCATCGCAACC
GCAAAGGCTACCGTTCACAACGAGGCCACAGCCGTGGCCGCAACCAGAACTCCCGCCGGCCATCCCGCGCCACGTGG
CTGTCCTTGTTCTCCAGTGAGGAGAGCAACTTGGGAGCCAACAACTATGATGACTACAGGATGGACTGGCTTGTGCC
TGCCACCTGTGAACCCATCCAGAGTGTCTTCTTCTTCTCTGGAGACAAGTACTACCGAGTCAATCTTCGCACACGGC
GAGTGGACACTGTGGACCCTCCCTACCCACGCTCCATCGCTCAGTACTGGCTGGGCTGCCCAGCTCCTGGCCATCTG
TAGGAGTCAGAGCCCACATGGCCGGGCCCTCTGTAGCTCCCTCCTCCCATCTCCTTCCCCCAGCCCAATAAAGGTCC
CTTAGCCCCGAAAAAAAAGCKATAAT

<210> SEQ ID NO 9

<211> Length : 4,755

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 9

>Z21368_PEA_1_T10

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGGATTCTT
CACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTA
CAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGAATGGAGAG
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGG
ATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATG
AAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATC
CCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATC
AAGATGTGGAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATC
AATGCCTTTGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAA
TGTCTACACCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATC
TTAACAACACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGG

TGGCGAGAATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAA
GCATGGATTTGATTATGCAAAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCC
CCACAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAA
ACACTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGC
TCCAGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAAT
ACTTACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATA
TGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTC
TCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTC
CTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATT
CCTAGTGGAAAGAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCA
AATATGAACGGGTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAA
TGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAG
CACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCA
GCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCAT
ACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATT
GCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCA
GTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACAC
AAGTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCA
TAAGGCATACATTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGA
AGAGAAGGAAGCCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAG
AAATTAAAGAGCCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAA
CAACCGTAGGAGGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCA
CTTGCTTCACGCATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCTGGGATCTTTCTGTGCTTGCACGAGT
TCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAATTTTCTTTTCTGTGAGTTTGCTACTGG
CTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAGCTCACAAATACAGTGCACACGGTAGAACGAGGCATTT
TGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAAT
CTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATC
AGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTAC
AGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAA
GAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGC
ACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTG
ACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAA
CGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTA
ATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATT
CGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGT
GCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGG
GGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCA
CTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAA

TAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGC
AAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGG
GTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGAC
ATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGT
AAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCG
TCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTT
ACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGA
TCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGC
AGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAAAATGA

<210> SEQ ID NO 10

<211> Length : 4,677

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 10

>Z21368_PEA_1_T11

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGGATTCTT
CACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTA
CAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGAATGGAGAG
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGG
ATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATG
AAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATC
CCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATC
AAGATGTGGAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATC
AATGCCTTTGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAA
TGTCTACACCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATC
TTAACAACACTGGCTACAGAACAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTC
TAAGAGAATGTATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCC
CACAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAA
CACTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCT
CCAGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATA

CTTACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATAT
GACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCT
CAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCC
TCAAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTC
CTAGTGGAAAGAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAA
ATATGAACGGGTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAAT
GCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGC
ACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAG
CAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATA
CTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATTG
CAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAG
TGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACA
AGTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCAT
AAGGCATACATTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAA
GAGAAGGAAGCCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGA
AATTAAAGAGCCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAAC
AACCGTAGGAGGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCAC
TTGCTTCACGCATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCTGGGATCTTTCTGTGCTTGCACGAGTT
CTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAATTTTCTTTTCTGTGAGTTTGCTACTGGC
TTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAGCTCACAAATACAGTGCACACGGTAGAACGAGGCATTTT
GAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATC
TTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATCA
GCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTACA
GACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAAG
AGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCA
CGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTGA
CCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAAC
GACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTAA
TCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTC
GCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTG
CATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGG
GAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCAC
TCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAAT
AAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCA
AGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGG
TTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGACA
TAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGTA
AATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGT

CATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTA
CAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGAT
CACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCA
GAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAAAATGA

<210> SEQ ID NO 11

<211> Length : 2,790

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 11

>Z21368_PEA_1_T23

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGGATTCTT
CACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTA
CAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGAATGGAGAG
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGG
ATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATG
AAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATC
CCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATC
AAGATGTGGAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATC
AATGCCTTTGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAA
TGTCTACACCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATC
TTAACAACACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGG
TGGCGAGAATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAA
GCATGGATTTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTA
AGAGAATGTATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCA
CAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAACA
CTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCC
AGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACT
TACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGA
CTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCA
ACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTC

AAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCT
AGTGGAAAGAGGGTAATTATTGGTTCCTGGGGTGCTTCTGGGAACCAGTCCTAGTGGGCAGCTTTCCCTGCTGAGTA
TTTTTTTTCTCCTTATTTTTGTTTACTAAGCATGCAGATTTCGTAAACCTAGTCACAAGATTGAATGGTTTGCTGCT
TATTCTGTAGTGGTCAATAGAGTAATAATTGCTGGATCAGAATTGTAAAGAATAACCCTCAAGTTGGTTAATTGGTA
CAAAAACACAGTTAGATAGAAGTTATAGAATTTGATAGTATAGTTGGGACATTATCGTTAACAATAATTTATGTATA
TCTTAAAATAGCTAGAAGTGAAGAATTGCAAAGTTCCCAACACAAGGAAAAGATAAATGAGATGATGAATATCCCAA
TTATCTTGATTTGATCATTACACATTGTAGACTGGTATCCATATATCACACGTACCCCCAAAATATGTATAATTGTG
ATATATCAATTTTTAAAATACCAAAAAAGCAAGAGAATGACGACTCCACATCCCCCAAAAAGAATAAATTCTCATAA
GCTTGGACCAAAGCCTTTATCATGGGTGTAGATTTACTGTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGG
ATTGAGTTTCTCTACCATCCTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCTTCTTCCTACT
TCCTACTTCTTTTCCTTGTGTGCCTTTCCTAGTTTTAAATAGATAAATGTATGCCATTGTAATTATTTCCATTGTCA
CTTCTGGGTTTCCCCTTTTGGTTCATTAATACCCATTGCCTTGTTTTTCTCTGTACATAAATTAGGAGAGAGAAAAT
ATTTGTATAATTTTTTTA

<210> SEQ ID NO 12

<211> Length : 3,069

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 12

>Z21368_PEA_1_T24

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGGATTCTT
CACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTA
CAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGAATGGAGAG
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGG
ATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATG
AAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATC
CCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATC
AAGATGTGGAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATC
AATGCCTTTGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAA
TGTCTACACCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATC
TTAACAACACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGG

TGGCGAGAATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAA
GCATGGATTTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTA
AGAGAATGTATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCA
CAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAACA
CTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCC
AGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACT
TACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGA
CTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCA
ACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTC
AAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTGTGTCATTGTTCCTCCTCTCAGCCAGCCCCAAATACACTGAGC
TCCAGCTGGTGCCCAGAGCCAGCCAGCAGCTGAAGACATGGAGGCAGAATATGCCTTGCCCACAAGGATCACCCCAA
GCTGAGCATTTCTCAGCTGCTTGTGAATAGCATATTGATGGAGATGCACTCATGGTCTGTGGGAAGTGAGAGGTGTT
TCTTTAAATAAGCTGTTAGCACAGATCCATTTGGAAAAACGTCCAGATGCCAAAAGTAAATATTATCATTTTGCTTT
CAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGGGTAATTATTGGTTCCTG
GGGTGCTTCTGGGAACCAGTCCTAGTGGGCAGCTTTCCCTGCTGAGTATTTTTTTTCTCCTTATTTTTGTTTACTAA
GCATGCAGATTTCGTAAACCTAGTCACAAGATTGAATGGTTTGCTGCTTATTCTGTAGTGGTCAATAGAGTAATAAT
TGCTGGATCAGAATTGTAAAGAATAACCCTCAAGTTGGTTAATTGGTACAAAAACACAGTTAGATAGAAGTTATAGA
ATTTGATAGTATAGTTGGGACATTATCGTTAACAATAATTTATGTATATCTTAAAATAGCTAGAAGTGAAGAATTGC
AAAGTTCCCAACACAAGGAAAAGATAAATGAGATGATGAATATCCCAATTATCTTGATTTGATCATTACACATTGTA
GACTGGTATCCATATATCACACGTACCCCCAAAATATGTATAATTGTGATATATCAATTTTTAAAATACCAAAAAAG
CAAGAGAATGACGACTCCACATCCCCCAAAAAGAATAAATTCTCATAAGCTTGGACCAAAGCCTTTATCATGGGTGT
AGATTTACTGTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTCTACCATCCTCCCCACGT
TCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCTTCTTCCTACTTCCTACTTCTTTTCCTTGTGTGCCTTTCC
TAGTTTTAAATAGATAAATGTATGCCATTGTAATTATTTCCATTGTCACTTCTGGGTTTCCCCTTTTGGTTCATTAA
TACCCATTGCCTTGTTTTTCTCTGTACATAAATTAGGAGAGAGAAAATATTTGTATAATTTTTTTA

<210> SEQ ID NO 13

<211> Length : 5,384

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 13

>Z21368_PEA_1_T5

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA

TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGTGCTGAC
GGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAATTGCAGAAA
ATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGGATACCTAA
TTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATGAAGTATTC
TTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGT
TCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATCAAGATGTG
GAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATCAATGCCTT
TGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAATGTCTACA
CCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATCTTAACAAC
ACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGGTGGCGAGA
ATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAAGCATGGAT
TTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTAAGAGAATG
TATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTC
TAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTA
TGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTG
ATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCAT
TTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATA
TTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCAACATTGAC
TTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCT
GGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAA
GAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGG
GTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAATGCATTGAGGA
TACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGCACGCGGAACC
TCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAA
AGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATACTCGGCAGAC
ACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGC
AACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAAC
AGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACAAGTGTTTTAT
TCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATAAGGCATACA
TTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAG
CCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAG
CCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGA
GGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACG
CATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCCTCACAAATACAGTGCACACGGTAGAACGAGGCATTT
TGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAAT
CTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATC
AGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTAC
AGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAA

GAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGC
ACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTG
ACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAA
CGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTA
ATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATT
CGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGT
GCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGG
GGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCA
CTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAA
TAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGC
AAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGG
GTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGAC
ATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGT
AAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCG
TCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTT
ACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGA
TCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGC
AGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAAAATGAAGATTGCCTGCTCTCTCTG
TGCCTAGCCTCAAAGCGTTCATCATACATCATACCTTTAAGATTGCTATATTTTGGGTTATTTTCTTGACAGGAGAA
AAAGATCTAAAGATCTTTTATTTTCATCTTTTTTGGTTTTCTTGGCATGACTAAGAAGCTTAAATGTTGATAAAATA
TGACTAGTTTTGAATTTACACCAAGAACTTCTCAATAAAAGAAAATCATGAATGCTCCACAATTTCAACATACCACA
AGAGAAGTTAATTTCTTAACATTGTGTTCTATGATTATTTGTAAGACCTTCACCAAGTTCTGATATCTTTTAAAGAC
ATAGTTCAAAATTGCTTTTGAAAATCTGTATTCTTGAAAATATCCTTGTTGTGTATTAGGTTTTTAAATACCAGCTA
AAGGATTACCTCACTGAGTCATCAGTACCCTCCTATTCAGCTCCCCAAGATGATGTGTTTTTGCTTACCCTAAGAGA
GGTTTTCTTCTTATTTTTAGATAATTCAAGTGCTTAGATAAATTATGTTTTCTTTAAGTGTTTATGGTAAACTCTTT
TAAAGAAAATTTAATATGTTATAGCTGAATCTTTTTGGTAACTTTAAATCTTTATCATAGACTCTGTACATATGTTC
AAATTAGCTGCTTGCCTGATGTGTGTATCATCGGTGGGATGACAGAACAAACATATTTATGATCATGAATAATGTGC
TTTGTAAAAAGATTTCAAGTTATTAGGAAGCATACTCTGTTTTTTAATCATGTATAATATTCCATGATACTTTTATA
GAACAATTCTGGCTTCAGGAAAGTCTAGAAGCAATATTTCTTCAAATAAAAGGTGTTTAAACTTTTTTCTG

<210> SEQ ID NO 14

<211> Length : 4,524

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 14

>Z21368_PEA_1_T6

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGTGCTGAC
GGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAATTGCAGAAA
ATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGGATACCTAA
TTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATGAAGTATTC
TTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGT
TCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATCAAGATGTG
GAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATCAATGCCTT
TGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAATGTCTACA
CCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATCTTAACAAC
ACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGGTGGCGAGA
ATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAAGCATGGAT
TTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTAAGAGAATG
TATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTC
TAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTA
TGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTG
ATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCAT
TTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATA
TTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCAACATTGAC
TTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCT
GGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAA
GAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGG
GTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAATGCATTGAGGA
TACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGCACGCGGAACC
TCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAA
AGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATACTCGGCAGAC
ACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGC
AACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAAC
AGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACAAGTGTTTTAT
TCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATAAGGCATACA
TTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAG
CCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAG
CCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGA
GGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACG
CATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCCTCACAAATACAGTGCACACGGTAGAACGAGGCATTT

TGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAAT
CTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATC
AGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTAC
AGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAA
GAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGC
ACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTG
ACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAA
CGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTA
ATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATT
CGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGT
GCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGG
GGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCA
CTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAA
TAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGC
AAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGG
GTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGAC
ATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGT
AAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCG
TCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTT
ACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGA
TCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGC
AGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAAAATGA

<210> SEQ ID NO 15

<211> Length : 4,454

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 15

>Z21368_PEA_1_T9

AGGTTACTTGACTGGGAGTTCTCAGACCTCCAGTTTCAGCCCTGCCCTCAGCCTCCAATCCGTAAGAGACACCCAGC
CCCAGCAATTGGATTGGGCAGCCCGTCTTGACACACCACTGTGCTGAGTGCTTGAGGACGTGTTTCAACAGATGGTT
GGGGTTAGTGTGTGTCATCACATTCGAGTGGGGATTAAGAGAAGGAAGGCTGCCTTGCTGGAGCTGTGTGGTCTTCT
CCAAGTGAGAGTCGCAGGCAATAGAACTACTTTGCTTTTGGAGGAAAAGGAGGAATTCATTTTCAGCAGACACAAGA
AAAGCAGTTTTTTTTTCAGGTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGG
TTCTTCAAGGCAGAATAATTGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGA

GATTGATTATTCAACCAGGATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAA
CATTGGACCAAATACAATGAAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCC
TCTGTTCGACTGTCAGATCCCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCAACATTATT
CTTGTGCTTACCGATGATCAAGATGTGGAGCTGGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACAT
CCCCCCTGGGTGGCGAGAATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCA
TCAAAGAAAAGCATGGATTTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTC
AAAATGTCTAAGAGAATGTATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGA
CTCAGCCCCACAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATA
TGGATAAACACTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGC
AAAAGGCTCCAGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCT
GGAGAATACTTACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCA
TGCCATATGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAG
ATCGTTCTCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAA
GTCTGTCCTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTG
ATACATTCCTAGTGGAAAGAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCAC
TTGCCCAAATATGAACGGGTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAA
GTGGCAATGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCC
GGCAGAGCACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTAC
CGTGCCAGCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATT
TGTCCATACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAG
AAGAATTGCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAG
GCTTCCAGTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGT
GACACACAAGTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGA
AGGACCATAAGGCATACATTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGA
CATCTGAAGAGAAGGAAGCCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAA
GCAAGAGAAATTAAAGAGCCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCA
AGGAGAACAACCGTAGGAGGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCT
GGCCTCACTTGCTTCACGCATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCTGGGATCTTTCTGTGCTTG
CACGAGTTCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAATTTTCTTTTCTGTGAGTTTG
CTACTGGCTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAGCTCACAAATACAGTGCACACGGTAGAACGA
GGCATTTTGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACC
TAAGAATCTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAG
GTTAATCAGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTA
TGAGTACAGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCAC
TGCTGAAGAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCT
GCTGAGCACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACC
TCATTTGACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAAT
GGAATAACGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAA

GAGACTAATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCA
GCCCATTCGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATC
TTCCTGTGCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGC
ATAGCGGGGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTC
CTCTTCACTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGT
AAACTAATAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCA
TGTGAGCAAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGA
TATTTGGGTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCG
TAGGGACATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACAC
CCCTGGTAAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTC
AATGCCGTCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTT
TATTTTTACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACA
GTATGGATCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATC
GATTAGCAGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAAAATGA

SEQ ID NO: 16

>H53626_PEA_1_T15

GTCCGGACAGGCCGAGATGACGCCGAGCCCCCTGTTGCTGCTCCTGCTGCCGCCGCTGCTGCTGGGGGCCTTCCCGC
CGGCCGCCGCCGCCCGAGGCCCCCCAAAGATGGCGGACAAGGTGGTCCCACGGCAGGTGGCCCGGCTGGGCCGCACT
GTGCGGCTGCAGTGCCCAGTGGAGGGGGACCCGCCGCCGCTGACCATGTGGACCAAGGATGGCCGCACCATCCACAG
CGGCTGGAGCCGCTTCCGCGTGCTGCCGCAGGGGCTGAAGGTGAAGCAGGTGGAGCGGGAGGATGCCGGCGTGTACG
TGTGCAAGGCCACCAACGGCTTCGGCAGCCTGAGCGTCAACTACACCCTCGTCGTGCTGGATGACATTAGCCCAGGG
AAGGAGAGCCTGGGGCCCGACAGCTCCTCTGGGGGTCAAGAGGACCCCGCCAGCCAGCAGTGGGCACGACCGCGCTT
CACACAGCCCTCCAAGATGAGGCGCCGGGTGATCGCACGGCCCGTGGGTAGCTCCGTGCGGCTCAAGTGCGTGGCCA
GCGGGCACCCTCGGCCCGACATCACGTGGATGAAGGACGACCAGGCCTTGACGCGCCCAGAGGCCGCTGAGCCCAGG
AAGAAGAAGTGGACACTGAGCCTGAAGAACCTGCGGCCGGAGGACAGCGGCAAATACACCTGCCGCGTGTCGAACCG
CGCGGGCGCCATCAACGCCACCTACAAGGTGGATGTGATCCAGCGGACCCGTTCCAAGCCCGTGCTCACAGGCACGC
ACCCCGTGAACACGACGGTGGACTTCGGGGGGACCACGTCCTTCCAGTGCAAGGTGCGCAGCGACGTGAAGCCGGTG
ATCCAGTGGCTGAAGCGCGTGGAGTACGGCGCCGAGGGCCGCCACAACTCCACCATCGATGTGGGCGGCCAGAAGTT
TGTGGTGCTGCCCACGGGTGACGTGTGGTCGCGGCCCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCGTGCCC
GCCAGGACGATGCGGGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTACAGCTTCCGCAGCGCCTTCCTCACC
GTGCTGCCAGGTGCGCGGCTGCCACGCCACGCCACACCATGCTGGTGCCCGGACCCGCCCCCTGGGCCCGGCGTCCC
ACCCACCGGGTGGGGCCCCACCCTTCCCTCCCGGGCCGTGCTGGCCAGGTCATCTGCCGAGGGAGGGCAGCCCAGGG
GCACCGTCTCCACAGCCCCTGGGATGGGTCTGGGGTGCTCTCCTGGTCTTTGTGTCGGCGTTCCCCTCCCTACCTCC
TTTCCTCTCGCTCTTGCAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTCGGCCACTAGCCTGCCGTG

GCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCATCCTGGGCACCCTGCTCCTGTGGCTTTGCCAGGCCCAGA
AGAAGCCGTGCACCCCCGCGCCTGCCCCTCCCCTGCCTGGGCACCGCCCGCCGGGGACGGCCCGCGACCGCAGCGGA
GACAAGGACCTTCCCTCGTTGGCCGCCCTCAGCGCTGGCCCTGGTGTGGGGCTGTGTGAGGAGCATGGGTCTCCGGC
AGCCCCCCAGCACTTACTGGGCCCAGGCCCAGTTGCTGGCCCTAAGTTGTACCCCAAACTCTACACAGACATCCACA
CACACACACACACACACTCTCACACACACTCACACGTGGAGGGCAAGGTCCACCAGCACATCCACTATCAGTGCTAG
ACGGCACCGTATCTGCAGTGGGCACGGGGGGGCCGGCCAGACAGGCAGACTGGGAGGATGGAGGACGGAGCTGCAGA
CGAAGGCAGGGGACCCATGGCGAGGAGGAATGGCCAGCACCCCAGGCAGTCTGTGTGTGAGGCATAGCCCCTGGACA
CACACACACAGACACACACACTACCTGGATGCATGTATGCACACACATGCGCGCACACGTGCTCCCTGAAGGCACAC
GTACGCACACACGCACATGCACAGATATGCCGCCTGGGCACACAGATAAGCTGCCCAAATGCACGCACACGCACAGA
GACATGCCAGAACATACAAGGACATGCTGCCTGAACATACACACGCACACCCATGCGCAGATGTGCTGCCTGGACAC
ACACACACACACGGATATGCTGTCTGGACGCACACACGTGCAGATATGGTATCCGGACACACACGTGCACAGATATG
CTGCCTGGACACACAGATAATGCTGCCTTGACACACACATGCACGGATATTGCCTGGACACACACACACACACGCGT
GCACAGATATGCTGTCTGGACACGCACACACATGCAGATATGCTGCCTGGACACACACTTCCAGACACACGTGCACA
GGCGCAGATATGCTGCCTGGACACACGCAGATATGCTGTCTAGTCACACACACACGCAGACATGCTGTCCGGACACA
CACACGCATGCACAGATATGCTGTCCGGACACACACACGCACGCAGATATGCTGCCTGGACACACACACAGATAATG
CTGCCTCAACACTCACACACGTGCAGATATTGCCTGGACACACACATGTGCACAGATATGCTGTCTGGACATGCACA
CACGTGCAGATATGCTGTCCGGATACACACGCACGCACACATGCAGATATGCTGCCTGGGCACACACTTCCGGACAC
ACATGCACACACAGGTGCAGATATGCTGCCTGGACACACGCAGACTGACGTGCTTTTGGGAGGGTGTGCCGTGAAGC
CTGCAGTACGTGTGCCGTGAGGCTCATAGTTGATGAGGGACTTTCCCTGCTCCACCGTCACTCCCCCAACTCTGCCC
GCCTCTGTCCCCGCCTCAGTCCCCGCCTCCATCCCCGCCTCTGTCCCCTGGCCTTGGCGGCTATTTTTGCCACCTGC
CTTGGGTGCCCAGGAGTCCCCTACTGCTGTGGGCTGGGGTTGGGGGCACAGCAGCCCCAAGCCTGAGAGGCTGGAGC
CCATGGCTAGTGGCTCATCCCCACTGCATTCTCCCCCTGACACAGAGAAGGGGCCTTGGTATTTATATTTAAGAAAT
GAAGATAATATTAATAATGATGGAAGGAAGACTGGGTTGCAGGGACTGTGGTCTCTCCTGGGGCCCGGGACCCGCCT
GGTCTTTCAGCCATGCTGATGACCACACCCCGTCCAGGCCAGACACCACCCCCCACCCCACTGTCGTGGTGGCCCCA
GATCTCTGTAATTTTATGTAGAGTTTGAGCTGAAGCCCCGTATATTTAATTTATTTTGTTAAACATGAAAGTGCATC
CTTTCCCTCCA

SEQ ID NO: 17

>H53626_PEA_1_T16

GTCCGGACAGGCCGAGATGACGCCGAGCCCCCTGTTGCTGCTCCTGCTGCCGCCGCTGCTGCTGGGGGCCTTCCCGC
CGGCCGCCGCCGCCCGAGGCCCCCCAAAGATGGCGGACAAGGTGGTCCCACGGCAGGTGGCCCGGCTGGGCCGCACT
GTGCGGCTGCAGTGCCCAGTGGAGGGGGACCCGCCGCCGCTGACCATGTGGACCAAGGATGGCCGCACCATCCACAG
CGGCTGGAGCCGCTTCCGCGTGCTGCCGCAGGGGCTGAAGGTGAAGCAGGTGGAGCGGGAGGATGCCGGCGTGTACG
TGTGCAAGGCCACCAACGGCTTCGGCAGCCTGAGCGTCAACTACACCCTCGTCGTGCTGGATGACATTAGCCCAGGG
AAGGAGAGCCTGGGGCCCGACAGCTCCTCTGGGGGTCAAGAGGACCCCGCCAGCCAGCAGTGGGCACGACCGCGCTT
CACACAGCCCTCCAAGATGAGGCGCCGGGTGATCGCACGGCCCGTGGGTAGCTCCGTGCGGCTCAAGTGCGTGGCCA
GCGGGCACCCTCGGCCCGACATCACGTGGATGAAGGACGACCAGGCCTTGACGCGCCCAGAGGCCGCTGAGCCCAGG
AAGAAGAAGTGGACACTGAGCCTGAAGAACCTGCGGCCGGAGGACAGCGGCAAATACACCTGCCGCGTGTCGAACCG

CGCGGGCGCCATCAACGCCACCTACAAGGTGGATGTGATCCAGCGGACCCGTTCCAAGCCCGTGCTCACAGGCACGC
ACCCCGTGAACACGACGGTGGACTTCGGGGGGACCACGTCCTTCCAGTGCAAGACCCAAAACCGCCAGGGCCACCTG
TGGCCTCCTCGTCCTCGGCCACTAGCCTGCCGTGGCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCATCCTG
GGCACCCTGCTCCTGTGGCTTTGCCAGGCCCAGAAGAAGCCGTGCACCCCCGCGCCTGCCCCTCCCCTGCCTGGGCA
CCGCCCGCCGGGGACGGCCCGCGACCGCAGCGGAGACAAGGACCTTCCCTCGTTGGCCGCCCTCAGCGCTGGCCCTG
GTGTGGGGCTGTGTGAGGAGCATGGGTCTCCGGCAGCCCCCCAGCACTTACTGGGCCCAGGCCCAGTTGCTGGCCCT
AAGTTGTACCCCAAACTCTACACAGACATCCACACACACACACACACACACTCTCACACACACTCACACGTGGAGGG
CAAGGTCCACCAGCACATCCACTATCAGTGCTAGACGGCACCGTATCTGCAGTGGGCACGGGGGGGCCGGCCAGACA
GGCAGACTGGGAGGATGGAGGACGGAGCTGCAGACGAAGGCAGGGGACCCATGGCGAGGAGGAATGGCCAGCACCCC
AGGCAGTCTGTGTGTGAGGCATAGCCCCTGGACACACACACACAGACACACACACTACCTGGATGCATGTATGCACA
CACATGCGCGCACACGTGCTCCCTGAAGGCACACGTACGCACACACGCACATGCACAGATATGCCGCCTGGGCACAC
AGATAAGCTGCCCAAATGCACGCACACGCACAGAGACATGCCAGAACATACAAGGACATGCTGCCTGAACATACACA
CGCACACCCATGCGCAGATGTGCTGCCTGGACACACACACACACACGGATATGCTGTCTGGACGCACACACGTGCAG
ATATGGTATCCGGACACACACGTGCACAGATATGCTGCCTGGACACACAGATAATGCTGCCTTGACACACACATGCA
CGGATATTGCCTGGACACACACACACACACGCGTGCACAGATATGCTGTCTGGACACGCACACACATGCAGATATGC
TGCCTGGACACACACTTCCAGACACACGTGCACAGGCGCAGATATGCTGCCTGGACACACGCAGATATGCTGTCTAG
TCACACACACACGCAGACATGCTGTCCGGACACACACACGCATGCACAGATATGCTGTCCGGACACACACACGCACG
CAGATATGCTGCCTGGACACACACACAGATAATGCTGCCTCAACACTCACACACGTGCAGATATTGCCTGGACACAC
ACATGTGCACAGATATGCTGTCTGGACATGCACACACGTGCAGATATGCTGTCCGGATACACACGCACGCACACATG
CAGATATGCTGCCTGGGCACACACTTCCGGACACACATGCACACACAGGTGCAGATATGCTGCCTGGACACACGCAG
ACTGACGTGCTTTTGGGAGGGTGTGCCGTGAAGCCTGCAGTACGTGTGCCGTGAGGCTCATAGTTGATGAGGGACTT
TCCCTGCTCCACCGTCACTCCCCCAACTCTGCCCGCCTCTGTCCCCGCCTCAGTCCCCGCCTCCATCCCCGCCTCTG
TCCCCTGGCCTTGGCGGCTATTTTTGCCACCTGCCTTGGGTGCCCAGGAGTCCCCTACTGCTGTGGGCTGGGGTTGG
GGGCACAGCAGCCCCAAGCCTGAGAGGCTGGAGCCCATGGCTAGTGGCTCATCCCCACTGCATTCTCCCCCTGACAC
AGAGAAGGGGCCTTGGTATTTATATTTAAGAAATGAAGATAATATTAATAATGATGGAAGGAAGACTGGGTTGCAGG
GACTGTGGTCTCTCCTGGGGCCCGGGACCCGCCTGGTCTTTCAGCCATGCTGATGACCACACCCCGTCCAGGCCAGA
CACCACCCCCCACCCCACTGTCGTGGTGGCCCCAGATCTCTGTAATTTTATGTAGAGTTTGAGCTGAAGCCCCGTAT
ATTTAATTTATTTTGTTAAACATGAAAGTGCATCCTTTCCCTCCA

SEQ ID NO: 18

>H53626_PEA_1_node_15

GCCCCCCAAAGATGGCGGACAAGGTGGTCCCACGGCAGGTGGCCCGGCTGGGCCGCACTGTGCGGCTGCAGTGCCCA
GTGGAGGGGGACCCGCCGCCGCTGACCATGTGGACCAAGGATGGCCGCACCATCCACAGCGGCTGGAGCCGCTTCCG
CGTGCTGCCGCAGGGGCTGAAGGTGAAGCAGGTGGAGCGGGAGGATGCCGGCGTGTACGTGTGCAAGGCCACCAACG
GCTTCGGCAGCCTGAGC

SEQ ID NO: 19

>H53626_PEA_1_node_22

CACGACCGCGCTTCACACAGCCCTCCAAGATGAGGCGCCGGGTGATCGCACGGCCCGTGGGTAGCTCCGTGCGGCTC
AAGTGCGTGGCCAGCGGGCACCCTCGGCCCGACATCACGTGGATGAAGGACGACCAGGCCTTGACGCGCCCAGAGGC
CGCTGAGCCCAGGAAGAAGAAGTGGACACTGAGCCTGAAGAACCTGCGGCCGGAGGACAGCGGCAAATACACCTGCC
GCGTGTCGAACCGCGCGGGCGCCATCAACGCCACCTACAAGGTGGATGTGATCC

<210> SEQ ID NO 20

<211> Length : 1,362

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 20

>HUMGRP5E_T4

CCAAAATCTATGGGCTGGGACAGCAAAGATGTGGCCTACGAAGAGAAAGGTCTGGAGAATCAGAAGGCCTTCAAATG
GTGGTTCCAAATCCCTCCAGCAAAGCCCATCCATCTTTAGAGCTCACCCGTCTCCAGCTACACCCCCCACCCCTCCC
GGCCCAGATCAGGCAGCGGGGTCGCCCTCTCCAGGACTCTCAAGGCAGCTAAGGCTGGAGGCGCCGGCGAGCCTGGA
GAGGGAGGAGTTCACTAAATTGTGTTGGATGGAAGGCGTCGAGGACCGGAGGAATTAATCCGATGTGGGGAAGGCGG
ACGGGGCTACGAGGAAAAAAGAGGGGGCAATGTACACTCAGCCTTTTCATCACTCGGCGGGGAGATGGATGGTTTTC
CGGACCGGGCGTCCCAGCGCCCCGGTTAGCTATAGGGAGACGTCAGAGCGCTCTGGTCCGCGATAGAAGAGCCCCCC
AGCCCCCCCGCCCGGGCTTCCATATAAAGTAGGGGCCCTAGTGGAGGCCGCAGCAGTAGCACCAGCGGCTGCGGCGG
CGGAGCTCCTCCGAGGTCCGGGTCACCAGTCTCTGCTCTTCCCAGCCTCTCCGGCGCGCTCCAAGGGCTTCCCGTCG
GGACCATGCGCGGCAGTGAGCTCCCGCTGGTCCTGCTGGCGCTGGTCCTCTGCCTGGCGCCCCGGGGGCGAGCGGTC
CCGCTGCCTGCGGGCGGAGGGACCGTGCTGACCAAGATGTACCCGCGCGGCAACCACTGGGCGGTGGGGCACTTAAT
GGGGAAAAAGAGCACAGGGGAGTCTTCTTCTGTTTCTGAGAGAGGGAGCCTGAAGCAGCAGCTGAGAGAGTACATCA
GGTGGGAAGAAGCTGCAAGGAATTTGCTGGGTCTCATAGAAGCAAAGGAGAACAGAAACCACCAGCCACCTCAACCC
AAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAGGATAGCAGCAACTTCAAAGATGTAGGTTCAAAAGGCAA
AGGTTCTCAACGTGAAGGAAGGAACCCCCAGCTGAACCAGCAATGATAATGATGGCCTCTCTCAAAAGAGAAAAACA
AAACCCCTAAGAGACTGCGTTCTGCAAGCATCAGTTCTACGGATCATCAACAAGATTTCCTTGTGCAAAATATTTGA
CTATTCTGTATCTTTCATCCTTGACTAAATTCGTGATTTTCAAGCAGCATCTTCTGGTTTAAACTTGTTTGCTGTGA
ACAATTGTCGAAAAGAGTCTTCCAATTAATGCTTTTTTATATCTAGGCTACCTGTTGGTTAGATTCAAGGCCCCGAG
CTGTTACCATTCACAATAAAAGCTTAAACACATTGTCCAAAGGGCAGGCTGTT

<210> SEQ ID NO 21

<211> Length : 1,376

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 21

>HUMGRP5E_T5

CCAAAATCTATGGGCTGGGACAGCAAAGATGTGGCCTACGAAGAGAAAGGTCTGGAGAATCAGAAGGCCTTCAAATG
GTGGTTCCAAATCCCTCCAGCAAAGCCCATCCATCTTTAGAGCTCACCCGTCTCCAGCTACACCCCCCACCCCTCCC
GGCCCAGATCAGGCAGCGGGGTCGCCCTCTCCAGGACTCTCAAGGCAGCTAAGGCTGGAGGCGCCGGCGAGCCTGGA
GAGGGAGGAGTTCACTAAATTGTGTTGGATGGAAGGCGTCGAGGACCGGAGGAATTAATCCGATGTGGGGAAGGCGG
ACGGGGCTACGAGGAAAAAAGAGGGGGCAATGTACACTCAGCCTTTTCATCACTCGGCGGGGAGATGGATGGTTTTC
CGGACCGGGCGTCCCAGCGCCCCGGTTAGCTATAGGGAGACGTCAGAGCGCTCTGGTCCGCGATAGAAGAGCCCCCC
AGCCCCCCCGCCCGGGCTTCCATATAAAGTAGGGGCCCTAGTGGAGGCCGCAGCAGTAGCACCAGCGGCTGCGGCGG
CGGAGCTCCTCCGAGGTCCGGGTCACCAGTCTCTGCTCTTCCCAGCCTCTCCGGCGCGCTCCAAGGGCTTCCCGTCG
GGACCATGCGCGGCAGTGAGCTCCCGCTGGTCCTGCTGGCGCTGGTCCTCTGCCTGGCGCCCCGGGGGCGAGCGGTC
CCGCTGCCTGCGGGCGGAGGGACCGTGCTGACCAAGATGTACCCGCGCGGCAACCACTGGGCGGTGGGGCACTTAAT
GGGGAAAAAGAGCACAGGGGAGTCTTCTTCTGTTTCTGAGAGAGGGAGCCTGAAGCAGCAGCTGAGAGAGTACATCA
GGTGGGAAGAAGCTGCAAGGAATTTGCTGGGTCTCATAGAAGCAAAGGAGAACAGAAACCACCAGCCACCTCAACCC
AAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAGGATAGCAGCAACTTCAAAGATGTAGGTTCAAAAGGCAA
AGACTCTCTGCTCCAGGTTCTCAACGTGAAGGAAGGAACCCCCAGCTGAACCAGCAATGATAATGATGGCCTCTCTC
AAAAGAGAAAAACAAAACCCCTAAGAGACTGCGTTCTGCAAGCATCAGTTCTACGGATCATCAACAAGATTTCCTTG
TGCAAAATATTTGACTATTCTGTATCTTTCATCCTTGACTAAATTCGTGATTTTCAAGCAGCATCTTCTGGTTTAAA
CTTGTTTGCTGTGAACAATTGTCGAAAAGAGTCTTCCAATTAATGCTTTTTTATATCTAGGCTACCTGTTGGTTAGA
TTCAAGGCCCCGAGCTGTTACCATTCACAATAAAAGCTTAAACACATTGTCCAAAGGGCAGGCTGTT

<210> SEQ ID NO 22

<211> Length : 902

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 22

>D56406_PEA_1_T3

TTCACTCACTTTCAAAGCCAGCTGAAGGAAAGAGGAAGTGCTAGAGAGAGCCCCCTTCAGTGTGCTTCTGACTTTTA
CGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGGAATGAAAATCCAGCTTGTATGCATGCTACTCCTGGCTT
TCAGCTCCTGGAGTCTGTGCTCAGATTCAGAAGAGGAAATGAAAGCATTAGAAGCAGATTTCTTGACCAATATGCAT
ACATCAAAGATTAGTAAAGCACATGTTCCCTCTTGGAAGATGACTCTGCTAAATGTTTGCAGTCTTGTAAATAATTT

GAACAGCCCAGCTGAGGAAACAGGAGAAGTTCATGAAGAGGAGCTTGTTGCAAGAAGGAAACTTCCTACTGCTTTAG
ATGGCTTTAGCTTGGAAGCAATGTTGACAATATACCAGCTCCACAAAATCTGTCACAGCAGGGCTTTTCAACACTGG
GAGGCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGACGGGCGGATCACGAGGTCAAGAGATGGA
GACCATCCCGGCTAACACGTTAATCCAGGAAGATATTCTTGATACTGGAAATGACAAAAATGGAAAGGAAGAAGTCA
TAAAGAGAAAAATTCCTTATATTCTGAAACGGCAGCTGTATGAGAATAAACCCAGAAGACCCTACATACTCAAAAGA
GATTCTTACTATTACTGAGAGAATAAATCATTTATTTACATGTGATTGTGATTCATCATCCCTTAATTAAATATCAA
ATTATATTTGTGTGAAAATGTGACAAACACACTTATCTGTCTCTTCTACAATTGTGGTTTATTGAATGTGATTTTTC
TGCACTAATATAAATTAGACTAAGTGTTTTCAAATAAATCTAAATCTTCAGCATG

<210> SEQ ID NO 23

<211> Length : 1,239

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 23

>D56406_PEA_1_T6

TTCACTCACTTTCAAAGCCAGCTGAAGGAAAGAGGAAGTGCTAGAGAGAGCCCCCTTCAGTGTGCTTCTGACTTTTA
CGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGGAATGAAAATCCAGCTTGTATGCATGCTACTCCTGGCTT
TCAGCTCCTGGAGTCTGTGCTCAGAAGAGGAAATGAAAGCATTAGAAGCAGATTTCTTGACCAATATGCATACATCA
AAGATTAGTAAAGCACATGTTCCCTCTTGGAAGATGACTCTGCTAAATGTTTGCAGTCTTGTAAATAATTTGAACAG
CCCAGCTGAGGAAACAGGAGAAGTTCATGAAGAGGAGCTTGTTGCAAGAAGGAAACTTCCTACTGCTTTAGATGGCT
TTAGCTTGGAAGCAATGTTGACAATATACCAGCTCCACAAAATCTGTCACAGCAGGGCTTTTCAACACTGGGAGTTA
ATCCAGGAAGATATTCTTGATACTGGAAATGACAAAAATGGAAAGGAAGAAGTCATAAAGAGAAAAATTCCTTATAT
TCTGAAACGGCAGCTGTATGAGAATAAACCCAGAAGACCCTACATACTCAAAAGAGATTCTTACTATTACTGAGAGA
ATAAATCATTTATTTACATGTGATTGTGATTCATCATCCCTTAATTAAATATCAAATTATATTTGTGTGAAAATGTG
ACAAACACACTTATCTGTCTCTTCTACAATTGTGGTTTATTGAATGTGATTTTTCTGCACTAATATAAATTAGACTA
AGTGTTTTCAAATAAATCTAAATCTTCAGCATGATGTGTTGTGTATAATTGGAGTAGATATTAATTAAGTCACCTGT
ATAATGTTTTGTAATTTTGCAAAACATATCTTGAGTTGTTTAAACAGTCAAAATGTTTGATATTTTATACCAGCTTA
TGAGCTCAAAGTACTACAGCAAAGCCTAGCCTGCATATCATTCACCCAAAACAAAGTAATAGCGCCTCTTTTATTAT
TTTGACTGAATGTTTTATGGAATTGAAAGAAACATACGTTCTTTTCAAGACTTCCTCATGAATCTCTCAATTATAGG
AAAAGTTATTGTGATAAAATAGGAACAGCTGAAAGATTGATTAATGAACTATTGTTAATTCTTCCTATTTTAATGAA
TGACATTGAACTGAATTTTTTGTCTGTTAAATGAACTTGATAGCTAATAAAAAGACAACTAGCCATCAAAATCAAAA
GTTTCTC

<210> SEQ ID NO 24

<211> Length : 1,020

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 24

>D56406_PEA_1_T7

TTCACTCACTTTCAAAGCCAGCTGAAGGAAAGAGGAAGTGCTAGAGAGAGCCCCCTTCAGTGTGCTTCTGACTTTTA
CGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGGAATGAAAATCCAGCTTGTATGCATGCTACTCCTGGCTT
TCAGCTCCTGGAGTCTGTGCTCAGATTCAGAAGAGGAAATGAAAGCATTAGAAGCAGATTTCTTGACCAATATGCAT
ACATCAAAGTTAATCCAGGAAGATATTCTTGATACTGGAAATGACAAAAATGGAAAGGAAGAAGTCATAAAGAGAAA
AATTCCTTATATTCTGAAACGGCAGCTGTATGAGAATAAACCCAGAAGACCCTACATACTCAAAAGAGATTCTTACT
ATTACTGAGAGAATAAATCATTTATTTACATGTGATTGTGATTCATCATCCCTTAATTAAATATCAAATTATATTTG
TGTGAAAATGTGACAAACACACTTATCTGTCTCTTCTACAATTGTGGTTTATTGAATGTGATTTTTCTGCACTAATA
TAAATTAGACTAAGTGTTTTCAAATAAATCTAAATCTTCAGCATGATGTGTTGTGTATAATTGGAGTAGATATTAAT
TAAGTCACCTGTATAATGTTTTGTAATTTTGCAAAACATATCTTGAGTTGTTTAAACAGTCAAAATGTTTGATATTT
TATACCAGCTTATGAGCTCAAAGTACTACAGCAAAGCCTAGCCTGCATATCATTCACCCAAAACAAAGTAATAGCGC
CTCTTTTATTATTTTGACTGAATGTTTTATGGAATTGAAAGAAACATACGTTCTTTTCAAGACTTCCTCATGAATCT
CTCAATTATAGGAAAAGTTATTGTGATAAAATAGGAACAGCTGAAAGATTGATTAATGAACTATTGTTAATTCTTCC
TATTTTAATGAATGACATTGAACTGAATTTTTTGTCTGTTAAATGAACTTGATAGCTAATAAAAAGACAACTAGCCA
TCAAAATCAAAAGTTTCTC

<210> SEQ ID NO 25

<211> Length : 1,737

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 25

>F05068_PEA_1_T3

AAGAAAGGGAAGGCAACCGGGCAGCCCAGGCCCCGCCCCGCCGCTCCCCCACCCGTGCGCTTATAAAGCACAGGAAC
CAGAGCTGGCCACTCAGTGGTTTCTTGGTGACACTGGATAGAACAGCTCAAGCCTTGCCACTTCGGGCTTCTCACTG
CAGCTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACTTTCTCCTTCAAGTACTTGGCAGAT
CACTCTCTTAGCAGGGTCTGCGCTTCGCAGCCGGGATGAAGCTGGTTTCCGTCGCCCTGATGTACCTGGGTTCGCTC
GCCTTCCTAGGCGCTGACACCGCTCGGTTGGATGTCGCGTCGGAGTTTCGAAAGAAGTGAGTCCGGGCAGCGCCTTC

CCCCTTGCTGGTACCTGGCAGGCAAGGGGAACTGACCGTTGGTCCCGAAGGTCTAGAAGTGAATGGGAGCAGGGACA
GGCCTGGGCGTCACCTGAACGCACGCGAATCGGGTCTGCTTGTGTTTTCCAGGTGGAATAAGTGGGCTCTGAGTCGT
GGGAAGAGGGAACTGCGGATGTCCAGCAGCTACCCCACCGGGCTCGCTGACGTGAAGGCCGGGCCTGCCCAGACCCT
TATTCGGCCCCAGGACATGAAGGGTGCCTCTCGAAGCCCCGAAGACAGCAGTCCGGATGCCGCCCGCATCCGAGTCA
AGCGCTACCGCCAGAGCATGAACAACTTCCAGGGCCTCCGGAGCTTTGGCTGCCGCTTCGGGACGTGCACGGTGCAG
AAGCTGGCACACCAGATCTACCAGTTCACAGATAAGGACAAGGACAACGTCGCCCCCAGGAGCAAGATCAGCCCCCA
GGGCTACGGCCGCCGGCGCCGGCGCTCCCTGCCCGAGGCCGGCCCGGGTCGGACTCTGGTGTCTTCTAAGCCACAAG
CACACGGGGCTCCAGCCCCCCCGAGTGGAAGTGCTCCCCACTTTCTTTAGGATTTAGGCGCCCATGGTACAAGGAAT
AGTCGCGCAAGCATCCCGCTGGTGCCTCCCGGGACGAAGGACTTCCCGAGCGGTGTGGGGACCGGGCTCTGACAGCC
CTGCGGAGACCCTGAGTCCGGGAGGCACCGTCCGGCGGCGAGCTCTGGCTTTGCAAGGGCCCCTCCTTCTGGGGGCT
TCGCTTCCTTAGCCTTGCTCAGGTGCAAGTGCCCCAGGGGGCGGGGTGCAGAAGAATCCGAGTGTTTGCCAGGCTTA
AGGAGAGGAGAAACTGAGAAATGAATGCTGAGACCCCCGGAGCAGGGGTCTGAGCCACAGCCGTGCTCGCCCACAAA
CTGATTTCTCACGGCGTGTCACCCCACCAGGGCGCAAGCCTCACTATTACTTGAACTTTCCAAAACCTAAAGAGGAA
AAGTGCAATGCGTGTTGTACATACAGAGGTAACTATCAATATTTAAGTTTGTTGCTGTCAAGATTTTTTTTGTAACT
TCAAATATAGAGATATTTTTGTACGTTATATATTGTATTAAGGGCATTTTAAAAGCAATTATATTGTCCTCCCCCTA
TTTTAAGACGTGAATGTCTCAGCGAGGTGTAAAGTTGTTCGCCGCGTGGAATGTGAGTGTGTTTGTGTGCATGAAAG
AGAAAGACTGATTACCTCCTGTGTGGAAGAAGGAAACACCGAGTCTCTGTATAATCTATTTACATAAAATGGGTGAT
ATGCGAACAGCAAACCAATAAACTGTCTCAATGCTGAATAAAA

<210> SEQ ID NO 26

<211> Length : 1,820

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 26

>F05068_PEA_1_T4

AAGAAAGGGAAGGCAACCGGGCAGCCCAGGCCCCGCCCCGCCGCTCCCCCACCCGTGCGCTTATAAAGCACAGGAAC
CAGAGCTGGCCACTCAGTGGTTTCTTGGTGACACTGGATAGAACAGCTCAAGCCTTGCCACTTCGGGCTTCTCACTG
CAGCTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACTTTCTCCTTCAAGTACTTGGCAGAT
CACTCTCTTAGCAGGGTCTGCGCTTCGCAGCCGGGATGAAGCTGGTTTCCGTCGCCCTGATGTACCTGGGTTCGCTC
GCCTTCCTAGGCGCTGACACCGCTCGGTTGGATGTCGCGTCGGAGTTTCGAAAGAAGTGGAATAAGTGGGCTCTGAG
TCGTGGGAAGAGGGAACTGCGGATGTCCAGCAGCTACCCCACCGGGCTCGCTGACGTGAAGGCCGGGCCTGCCCAGA
CCCTTATTCGGCCCCAGGACATGAAGGGTGCCTCTCGAAGCCCCGAAGACAGGTAACTACGCCCTGTGCTGTCCAGG
GACGGGAGGGAAGGAAGGTGTGCGGGAGGAGTTCTCTGTCTCCACTCCCCTGGCCCGGGGGATCGTCGGGGCTGGAC
CGCAGCTCAGATGGCGCGAGCAGTTTCCAGCTCCCTCTGGCTCTAGAATGGCTCCCGTTCCCGGTGTTGGGGCCAAA
GCTCTGCTTGATGGGGTCTCAAGTTGCCTTTCTTCCCCCTCCCCCCGCCCGCAGCAGTCCGGATGCCGCCCGCATCC

GAGTCAAGCGCTACCGCCAGAGCATGAACAACTTCCAGGGCCTCCGGAGCTTTGGCTGCCGCTTCGGGACGTGCACG
GTGCAGAAGCTGGCACACCAGATCTACCAGTTCACAGATAAGGACAAGGACAACGTCGCCCCCAGGAGCAAGATCAG
CCCCCAGGGCTACGGCCGCCGGCGCCGGCGCTCCCTGCCCGAGGCCGGCCCGGGTCGGACTCTGGTGTCTTCTAAGC
CACAAGCACACGGGGCTCCAGCCCCCCCGAGTGGAAGTGCTCCCCACTTTCTTTAGGATTTAGGCGCCCATGGTACA
AGGAATAGTCGCGCAAGCATCCCGCTGGTGCCTCCCGGGACGAAGGACTTCCCGAGCGGTGTGGGGACCGGGCTCTG
ACAGCCCTGCGGAGACCCTGAGTCCGGGAGGCACCGTCCGGCGGCGAGCTCTGGCTTTGCAAGGGCCCCTCCTTCTG
GGGGCTTCGCTTCCTTAGCCTTGCTCAGGTGCAAGTGCCCCAGGGGGCGGGGTGCAGAAGAATCCGAGTGTTTGCCA
GGCTTAAGGAGAGGAGAAACTGAGAAATGAATGCTGAGACCCCCGGAGCAGGGGTCTGAGCCACAGCCGTGCTCGCC
CACAAACTGATTTCTCACGGCGTGTCACCCCACCAGGGCGCAAGCCTCACTATTACTTGAACTTTCCAAAACCTAAA
GAGGAAAAGTGCAATGCGTGTTGTACATACAGAGGTAACTATCAATATTTAAGTTTGTTGCTGTCAAGATTTTTTTT
GTAACTTCAAATATAGAGATATTTTTGTACGTTATATATTGTATTAAGGGCATTTTAAAAGCAATTATATTGTCCTC
CCCCTATTTTAAGACGTGAATGTCTCAGCGAGGTGTAAAGTTGTTCGCCGCGTGGAATGTGAGTGTGTTTGTGTGCA
TGAAAGAGAAAGACTGATTACCTCCTGTGTGGAAGAAGGAAACACCGAGTCTCTGTATAATCTATTTACATAAAATG
GGTGATATGCGAACAGCAAACCAATAAACTGTCTCAATGCTGAATAAAA

<210> SEQ ID NO 27

<211> Length : 1,970

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 27

>F05068_PEA_1_T6

AAGAAAGGGAAGGCAACCGGGCAGCCCAGGCCCCGCCCCGCCGCTCCCCCACCCGTGCGCTTATAAAGCACAGGAAC
CAGAGCTGGCCACTCAGTGGTTTCTTGGTGACACTGGATAGAACAGCTCAAGCCTTGCCACTTCGGGCTTCTCACTG
CAGCTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACTTTCTCCTTCAAGTACTTGGCAGAT
CACTCTCTTAGCAGGGTCTGCGCTTCGCAGCCGGGATGAAGCTGGTTTCCGTCGCCCTGATGTACCTGGGTTCGCTC
GCCTTCCTAGGCGCTGACACCGCTCGGTTGGATGTCGCGTCGGAGTTTCGAAAGAAGTGAGTCCGGGCAGCGCCTTC
CCCCTTGCTGGTACCTGGCAGGCAAGGGGAACTGACCGTTGGTCCCGAAGGTCTAGAAGTGAATGGGAGCAGGGACA
GGCCTGGGCGTCACCTGAACGCACGCGAATCGGGTCTGCTTGTGTTTTCCAGGTGGAATAAGTGGGCTCTGAGTCGT
GGGAAGAGGGAACTGCGGATGTCCAGCAGCTACCCCACCGGGCTCGCTGACGTGAAGGCCGGGCCTGCCCAGACCCT
TATTCGGCCCCAGGACATGAAGGGTGCCTCTCGAAGCCCCGAAGACAGGTAACTACGCCCTGTGCTGTCCAGGGACG
GGAGGGAAGGAAGGTGTGCGGGAGGAGTTCTCTGTCTCCACTCCCCTGGCCCGGGGGATCGTCGGGGCTGGACCGCA
GCTCAGATGGCGCGAGCAGTTTCCAGCTCCCTCTGGCTCTAGAATGGCTCCCGTTCCCGGTGTTGGGGCCAAAGCTC
TGCTTGATGGGGTCTCAAGTTGCCTTTCTTCCCCCTCCCCCCGCCCGCAGCAGTCCGGATGCCGCCCGCATCCGAGT
CAAGCGCTACCGCCAGAGCATGAACAACTTCCAGGGCCTCCGGAGCTTTGGCTGCCGCTTCGGGACGTGCACGGTGC
AGAAGCTGGCACACCAGATCTACCAGTTCACAGATAAGGACAAGGACAACGTCGCCCCCAGGAGCAAGATCAGCCCC

CAGGGCTACGGCCGCCGGCGCCGGCGCTCCCTGCCCGAGGCCGGCCCGGGTCGGACTCTGGTGTCTTCTAAGCCACA
AGCACACGGGGCTCCAGCCCCCCCGAGTGGAAGTGCTCCCCACTTTCTTTAGGATTTAGGCGCCCATGGTACAAGGA
ATAGTCGCGCAAGCATCCCGCTGGTGCCTCCCGGGACGAAGGACTTCCCGAGCGGTGTGGGGACCGGGCTCTGACAG
CCCTGCGGAGACCCTGAGTCCGGGAGGCACCGTCCGGCGGCGAGCTCTGGCTTTGCAAGGGCCCCTCCTTCTGGGGG
CTTCGCTTCCTTAGCCTTGCTCAGGTGCAAGTGCCCCAGGGGGCGGGGTGCAGAAGAATCCGAGTGTTTGCCAGGCT
TAAGGAGAGGAGAAACTGAGAAATGAATGCTGAGACCCCCGGAGCAGGGGTCTGAGCCACAGCCGTGCTCGCCCACA
AACTGATTTCTCACGGCGTGTCACCCCACCAGGGCGCAAGCCTCACTATTACTTGAACTTTCCAAAACCTAAAGAGG
AAAAGTGCAATGCGTGTTGTACATACAGAGGTAACTATCAATATTTAAGTTTGTTGCTGTCAAGATTTTTTTTGTAA
CTTCAAATATAGAGATATTTTTGTACGTTATATATTGTATTAAGGGCATTTTAAAAGCAATTATATTGTCCTCCCCC
TATTTTAAGACGTGAATGTCTCAGCGAGGTGTAAAGTTGTTCGCCGCGTGGAATGTGAGTGTGTTTGTGTGCATGAA
AGAGAAAGACTGATTACCTCCTGTGTGGAAGAAGGAAACACCGAGTCTCTGTATAATCTATTTACATAAAATGGGTG
ATATGCGAACAGCAAACCAATAAACTGTCTCAATGCTGAATAAAA

<210> SEQ ID NO 28

<211> Length : 1,745

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 28

>H14624_T20

TTATGCTCCCGCGGAGGCCAAGCGGACTCCCTGACAGGACAGAATCTGAACGTGAGAGTGAAGGTCTTGCCTGTCCA
GAAACTCTTGTAGCCAGCACAGGTTTAAACAAGAAGCCAAATTGTTCTGGAGAGATTGCTGGGGGCTTTCTTTGTGC
CTCAAGCTTCTTCAGTGCCCTGAGCACAGGAAACACTCAAGCAGAGAAGCAGAGCCAAACCCAGGATACGGGAGGTC
GAGGCTCTTCCGTAGACCTGCAGCATTGGGGTGGGATGATGTTCATTCTGTGTGTGTTCTGGACCAAGCCCCTCTCC
AGGGACCTATGGGCAGCCCCCTTTAAGCAAGATGCCCGGTGGAGTGGGCATCCACCATCACTTACCCTGGGCTTGGG
TGAATAGATTTTCCGTGCCTTAAATGGGCAGGGAGGGGGTAAACATGGACGGTCCATTGGTACAAATAAAAGCCTTT
GGTGGGTTTTGATCAATTGCAAGGATCGAAGGAGACCTGTGGACCTGAGGTCAACTGGCAGCAGAGAAGAGTCTGGG
TTCGTGAAGGCGCCGCCGCGGTGCCGCGCCACGTATTTGCATAAAAAAGGCCAAGAAAACTCTGGCTGTGCCCCAGC
AACGGCTCATTCTGCTCCCCCGGGTCGGAGCCCCCCGGAGCTGCGCGCGGGCTTGCAGCGCCTCGCCCGCGCTGTCC
TCCCGGTGTCCCGCTTCTCCGCGCCCCAGCCGCCGGCTGCCAGCTTTTCGGGGCCCCGAGTCGCACCCAGCGAAGAG
AGCGGGCCCGGGACAAGCTCGAACTCCGGCCGCCTCGCCCTTCCCCGGCTCCGCTCCCTCTGCCCCCTCGGGGTCGC
GCGCCCACGATGCTGCAGGGCCCTGGCTCGCTGCTGCTGCTCTTCCTCGCCTCGCACTGCTGCCTGGGCTCGGCGCG
CGGGCTCTTCCTCTTTGGCCAGCCCGACTTCTCCTACAAGCGCAGCAATTGCAAGCCCATCCCGGCCAACCTGCAGC
TGTGCCACGGCATCGAATACCAGAACATGCGGCTGCCCAACCTGCTGGGCCACGAGACCATGAAGGAGGTGCTGGAG
CAGGCCGGCGCTTGGATCCCGCTGGTCATGAAGCAGTGCCACCCGGACACCAAGAAGTTCCTGTGCTCGCTCTTCGC

CCCCGTCTGCCTCGATGACCTAGACGAGACCATCCAGCCATGCCACTCGCTCTGCGTGCAGGTGAAGGACCGCTGCG
CCCCGGTCATGTCCGCCTTCGGCTTCCCCTGGCCCGACATGCTTGAGTGCGACCGTTTCCCCCAGGACAACGACCTT
TGCATCCCCCTCGCTAGCAGCGACCACCTCCTGCCAGCCACCGAGGAAGGTAAGCCTTCCCTCTTGCTTCCCCACTC
CCTGCTGGGCTGAGACGCTCCCAGGAGATCCCGCCCCTGCCACGCATCCCAGTGCATCCCTGCTTGGGGTGCCAGTA
GCGGGAAGGGCAGAAGTTCTGCCTGACCTGGTCTGTCATCACAACAAGCCTGTATCAAATTTGAGGCACCCCTCCCA
CGCCGCCCAAGTCTCGCGCATTCTCCTCCCGAGTTGTACCAGCTATACTTAAGGGCAGTTTAAAAATAAAACAAACA
AACAAAAACAACAAAACTAAAAAAACGAAGAACTGAACGGCGGTTTAAAAAAAAATAGATACACGATTATTGTTAAA
GATGCTAGCACTGGAGCTGCGCAGAGCGTTGGAAGTGGTGTTTGGTGGAGG

<210> SEQ ID NO 29

<211> Length : 3,170

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 29

>H38804_PEA_1_T24

GACTGGGTTGACCGATGCTGGGCAGCTGAGCGGACCAATCGGCCCCCTAGACTGAGACGTTGGCGTTTGAAATCAGC
CAATGGCAGGTCTACACTGGAGCTTCCTCTCCGCCTCCTTCGCCTAGCCTGCGAGTGTTCTGAGGGAAGCAAGGAGG
CGGCGGCGGCCGCAGCGAGTGGCGAGTAGTGGAAACGTTGCTTCTGAGGGGAGCCCAAGGTAGGGAGGCGAGGCGAC
GGTGTGCGGGAGCGGGCTCTCCAGGGACTTCCCGGGTCCGCAACTGGCAGGGCCGTTCGATTCGCAGGGGATCCCGT
TTCGTTTCTGTTGTTTTCCCTTTATTTTTAGGAGTGCCCGGGGCGACGGGACCCCGGGAGAGGGGAAAGGGAACAGT
CTGGGGTCCGGGCATCGCTGTGGGCCGGGCTGGGTTTAGGGGGACGGCGGTGCGGGCTGGGCCGGTTTGGGCGCGGC
GGGGGCCGGATGATGGGGCGAGTCCGGACCTTGGCGGGCGAGTGCTCGGCGCAGGCGCAAGCGCAGAGTCTCCTCGC
GGTCGTCCTCTCGGCCCCTCCCTCTGGGGGGACCCCCAGTGCCAGGCTGTCAGTGCGCAGCCCCAGCCCGCGGGACC
CCTGGGGACTCTGGGCGCCTGTTCTGCAGATGACCGGTTCTAACGAGTTCAAGCTGAACCAGCCACCCGAGGATGGC
ATCTCCTCCGTGAAGTTCAGCCCCAACACCTCCCAGTTCCTGCTTGTCTCCTCCTGGGACACGTCCGTGCGTCTCTA
CGATGTGCCGGCCAACTCCATGCGGCTCAAGTACCAGCACACCGGCGCCGTCCTGGACTGCGCCTTCTACGATCCAA
CGCATGCCTGGAGTGGAGGACTAGATCATCAATTGAAAATGCATGATTTGAACACTGATCAAGAAAATCTTGTTGGG
ACCCATGATGCCCCTATCAGATGTGTTGAATACTGTCCAGAAGTGAATGTGATGGTCACTGGAAGTTGGGATCAGAC
AGTTAAACTGTGGGATCCCAGAACTCCTTGTAATGCTGGGACCTTCTCTCAGCCTGAAAAGGTATATACCCTCTCAG
TGTCTGGAGACCGGCTGATTGTGGGAACAGCAGGCCGCAGAGTGTTGGTGTGGGACTTACGGAACATGGGTTACGTG
CAGCAGCGCAGGGAGTCCAGCCTGAAATACCAGACTCGCTGCATACGAGCGTTTCCAAACAAGCAGGGTTATGTATT
AAGCTCTATTGAAGGCCGAGTGGCAGTTGAGTATTTGGACCCAAGCCCTGAGGTACAGAAGAAGAAGTATGCCTTCA
AATGTCACAGACTAAAAGAAAATAATATTGAGCAGATTTACCCAGTCAATGCCATTTCTTTTCACAATATCCACAAT

ACATTTGCCACAGGTGGTTCTGATGGCTTTGTAAATATTTGGGATCCATTTAACAAAAAGCGACTGTGCCAATTCCA
TCGGTACCCCACGAGCATCGCATCACTTGCCTTCAGTAATGATGGGACTACGCTTGCAATAGCGTCATCATATATGT
ATGAAATGGATGACACAGAACATCCTGAAGATGGTATCTTCATTCGCCAAGTGACAGATGCAGAAACAAAACCCAAG
TCACCATGTACTTGACAAGATTTCATTTACTTAAGTGCCATGTTGATGATAATAAAACAATTCGTACTCCCCAATGG
TGGATTTATTACTATTAAAGAAACCAGGGAAAATATTAATTTTAATATTATAACAACCTGAAAATAATGGAAAAGAG
GTTTTTGAATTTTTTTTTTTAAATAAACACCTTCTTAAGTGCATGAGATGGTTTGATGGTTTGCTGCATTAAAGGTA
TTTGGGCAAACAAAATTGGAGGGCAAGTGACTGCAGTTTTGAGAATCAGTTTTGACCTTGATGATTTTTTGTTTCCA
CTGTGGAAATAAATGTTTGTAAATAAGTGTAATAAAAATCCCTTTGCATTCTTTCTGGACCTTAAATGGTAGAGGAA
AAGGCTCGTGAGCCATTTGTTTCTTTTGCTGGTTATAGTTGCTAATTCTAAAGCTGCTTCAGACTGCTTCATGAGGA
GGTTAATCTACAATTAAACAATATTTCCTCTTGGCCGTCCATTATTTTCTGAAGCAGATGGTTCATCATTTCCTGGG
CTGTTAAACAAAGCGAGGTTAAGGTTAGACTCTTGGGAATCAGCTAGTTTTCAATCTTATTAGGGTGCAGAAGGAAA
ACTAATAAGAAAACCTCCTAATATCATTTTGTGACTGTAAACAATTATTTATTAGCAAACAATTGATCCCAGAAGGG
CAAATTGTTTGAGTCAGTAATGAGCTGAGAAAAGACAGAGCATATCTGTGTATTTGGAAAAATAATTGTAACGTAAT
TGCAGTGCATTTAGACAGGCATCTATTTGGACCTGTTTCTATCTCTAAATGAATTTTTGGAAACATTAATGAGGTTT
ACATATTTCTCTGACATTTATATAGTTCTTATGTCCATTTCAGTTGACCAGCCGCTGGTGATTAAAGTTAAAAAGAA
AAAAATTATAGTGAGAATGAGATTCATTTCAATGTAATGCACTAAAGCAGAACACGAACTTAGCTTGGCCTATTCTA
GGTAGTTCCAAATAGTATTTTTGTTGTCAAACTTTAAAATTTATATTAATTTGCAAATGTATGTCTCTGAGTAGGAC
TTGGACCTTTCCTGAGATTTATTTTATCCGTGATGTATTTTTTTTAATTCTTTTGATACAGAGAAGGGTCTTTTTTT
TTTTTAAGTATTTCAGTGAAAACTTGGTGTAAGTCTGAACCCATCTTTTGAAATGTATTTTCTTCATTGCAGGTCCA
CCTAATCATCCTGTGAAAGTGGTTTCTCTATGGAAAGCTTTGTTTGCTTCCTACAAATACATGCTTATTCCTTAAGG
GATGTGTTAGAGTTACTGTGGATTTCTCTGTTTTCTGTCTTACAAGAAACTTGTCTATGTACCTTAATACTTTGTTT
AGGATGAGGAGTCTTTGTGTCCCTGTACAGTAGTCTGACGTATTTCCCCTTCTGTCCCCTAGTAAGCCCAGTTGCTG
TATCTGAACAGTTTGAGCTCTTTTTGTAATATACTCTAAACCTGTTATTTCTGTGCTAATAAACGAGATGCAGAACC
CTTGAAAAATGGA

<210> SEQ ID NO 30

<211> Length : 4,161

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 30

>H38804_PEA_1_T8

GACTGGGTTGACCGATGCTGGGCAGCTGAGCGGACCAATCGGCCCCCTAGACTGAGACGTTGGCGTTTGAAATCAGC
CAATGGCAGGTCTACACTGGAGCTTCCTCTCCGCCTCCTTCGCCTAGCCTGCGAGTGTTCTGAGGGAAGCAAGGAGG
CGGCGGCGGCCGCAGCGAGTGGCGAGTAGTGGAAACGTTGCTTCTGAGGGGAGCCCAAGGTAGGGAGGCGAGGCGAC
GGTGTGCGGGAGCGGGCTCTCCAGGGACTTCCCGGGTCCGCAACTGGCAGGGCCGTTCGATTCGCAGGGGATCCCGT

TTCGTTTCTGTTGTTTTCCCTTTATTTTTAGGAGTGCCCGGGGCGACGGGACCCCGGGAGAGGGGAAAGGGAACAGT
CTGGGGTCCGGGCATCGCTGTGGGCCGGGCTGGGTTTAGGGGGACGGCGGTGCGGGCTGGGCCGGTTTGGGCGCGGC
GGGGGCCGGATGATGGGGCGAGTCCGGACCTTGGCGGGCGAGTGCTCGGCGCAGGCGCAAGCGCAGAGTCTCCTCGC
GGTCGTCCTCTCGGCCCCTCCCTCTGGGGGGACCCCCAGTGCCAGGCTGTCAGTGCGCAGCCCCAGCCCGCGGGACC
CCTGGGGACTCTGGGCGCCTGTTCTGCAGATGACCGGTTCTAACGAGTTCAAGCTGAACCAGCCACCCGAGGATGGC
ATCTCCTCCGTGAAGTTCAGCCCCAACACCTCCCAGTTCCTGCTTGTCTCCTCCTGGGACACGTCCGTGCGTCTCTA
CGATGTGCCGGCCAACTCCATGCGGCTCAAGTACCAGCACACCGGCGCCGTCCTGGACTGCGCCTTCTACGATCCAA
CGCATGCCTGGAGTGGAGGACTAGATCATCAATTGAAAATGCATGATTTGAACACTGATCAAGAAAATCTTGTTGGG
ACCCATGATGCCCCTATCAGATGTGTTGAATACTGTCCAGAAGTGAATGTGATGGTCACTGGAAGTTGGGATCAGAC
AGTTAAACTGTGGGATCCCAGAACTCCTTGTAATGCTGGGACCTTCTCTCAGCCTGAAAAGGTATATACCCTCTCAG
TGTCTGGAGACCGGCTGATTGTGGGAACAGCAGGCCGCAGAGTGTTGGTGTGGGACTTACGGAACATGGGTTACGTG
CAGCAGCGCAGGGAGTCCAGCCTGAAATACCAGACTCGCTGCATACGAGCGTTTCCAAACAAGCAGGGTTATGTATT
AAGCTCTATTGAAGGCCGAGTGGCAGTTGAGTATTTGGACCCAAGCCCTGAGGTACAGAAGAAGAAGTATGCCTTCA
AATGTCACAGACTAAAAGAAAATAATATTGAGCAGATTTACCCAGTCAATGCCATTTCTTTTCACAATATCCACAAT
ACATTTGCCACAGGTGGTTCTGATGGCTTTGTAAATATTTGGGATCCATTTAACAAAAAGCGACTGTGCCAATTCCA
TCGGTACCCCACGAGCATCGCATCACTTGCCTTCAGTAATGATGGGACTACGCTTGCAATAGCGTCATCATATATGT
ATGAAATGGATGACACAGAACATCCTGAAGATGGTATCTTCATTCGCCAAGTGACAGATGCAGAAACAAAACCCAAG
TGAGTATGCTTCACCTGTATTTGAGCCTTTTCTTGCATTCAACCCAGGATTTATTAATTTTTCTAAATTCATGAATA
GCATTGTTGATGCCTGCTCGATATTACAGCTGACTGTAGGGTTGGAGTTGATGTTATCATGTTCTCCCAAGCTTTCA
ATATCCGTAGGTTGATAGACGTCTGATGGATAAAATTGTGCCTAGTTGTTTTGTAGAGAAGAATGTCAAACTCTTAT
TCTTCTTGAATAGGCTCTATTATTTGAATCTCTGGAGTTATTACCAGCTCATTGCTTCAAAATTAAGTTGAGGAATT
CAAGAATAATTTATTTTAGTAAATTCTATTTAAGATGTTTAAGAATTTGAACTGCCAAAAATCTTTCCTCTCCACAG
AGGTTGTTTCTTTAATATTAACACAAAGTAAGTGACCTTCAGGTCTTATTGGAAACTCAGAGTAATATGGCCTTGCC
TGGAATTGCAAATTTCCTTAGTTTTGAAATTTTCATAGATGTCTTTGGTTCTTGGTTGTAACTGTTGACTGAGAAGA
GCCATTTACATTTTTTGATACCAACAGGGCAAAGCTTTTTACTTAATTACCTCTACCAGGCTTTAAGGGAAATCTGA
TACTTCAGCATGTGTTAAACTATAAAATACCTACTCCAAGTATCTGCCCAGTTCCTTGTCCCCTCTCCCCAGGCCCT
TAAAGGAAGTTCTCGATACATATTTGTAGAATAACTGAATGTTTTCAGGATTCCTGTACTTTGCTGAGTTAAAATGG
ATATGGTACCCTTGCTGATTGGTTGAGCCCCTAAGAGGGGGCAGAATATTAAATATTCCATATCAGATATGCTTTTA
CAGGTTTGACTTTAGAAAAGTCTTAGCATGTGAAGCCTGTTGGATAAAGGGCTGTGTTTGCATTTAATCTGTCACTT
TTGTATCTCCTGTCCTGGCTGGCCATTTTGATCTCATGCTGTTCTTTTTTTCTTTTGAACTTGTAGGTCACCATGTA
CTTGACAAGATTTCATTTACTTAAGTGCCATGTTGATGATAATAAAACAATTCGTACTCCCCAATGGTGGATTTATT
ACTATTAAAGAAACCAGGGAAAATATTAATTTTAATATTATAACAACCTGAAAATAATGGAAAAGAGGTTTTTGAAT
TTTTTTTTTTAAATAAACACCTTCTTAAGTGCATGAGATGGTTTGATGGTTTGCTGCATTAAAGGTATTTGGGCAAA
CAAAATTGGAGGGCAAGTGACTGCAGTTTTGAGAATCAGTTTTGACCTTGATGATTTTTTGTTTCCACTGTGGAAAT
AAATGTTTGTAAATAAGTGTAATAAAAATCCCTTTGCATTCTTTCTGGACCTTAAATGGTAGAGGAAAAGGCTCGTG
AGCCATTTGTTTCTTTTGCTGGTTATAGTTGCTAATTCTAAAGCTGCTTCAGACTGCTTCATGAGGAGGTTAATCTA
CAATTAAACAATATTTCCTCTTGGCCGTCCATTATTTTCTGAAGCAGATGGTTCATCATTTCCTGGGCTGTTAAACA
AAGCGAGGTTAAGGTTAGACTCTTGGGAATCAGCTAGTTTTCAATCTTATTAGGGTGCAGAAGGAAAACTAATAAGA

AAACCTCCTAATATCATTTTGTGACTGTAAACAATTATTTATTAGCAAACAATTGATCCCAGAAGGGCAAATTGTTT
GAGTCAGTAATGAGCTGAGAAAAGACAGAGCATATCTGTGTATTTGGAAAAATAATTGTAACGTAATTGCAGTGCAT
TTAGACAGGCATCTATTTGGACCTGTTTCTATCTCTAAATGAATTTTTGGAAACATTAATGAGGTTTACATATTTCT
CTGACATTTATATAGTTCTTATGTCCATTTCAGTTGACCAGCCGCTGGTGATTAAAGTTAAAAAGAAAAAAATTATA
GTGAGAATGAGATTCATTTCAATGTAATGCACTAAAGCAGAACACGAACTTAGCTTGGCCTATTCTAGGTAGTTCCA
AATAGTATTTTTGTTGTCAAACTTTAAAATTTATATTAATTTGCAAATGTATGTCTCTGAGTAGGACTTGGACCTTT
CCTGAGATTTATTTTATCCGTGATGTATTTTTTTTAATTCTTTTGATACAGAGAAGGGTCTTTTTTTTTTTTAAGTA
TTTCAGTGAAAACTTGGTGTAAGTCTGAACCCATCTTTTGAAATGTATTTTCTTCATTGCAGGTCCACCTAATCATC
CTGTGAAAGTGGTTTCTCTATGGAAAGCTTTGTTTGCTTCCTACAAATACATGCTTATTCCTTAAGGGATGTGTTAG
AGTTACTGTGGATTTCTCTGTTTTCTGTCTTACAAGAAACTTGTCTATGTACCTTAATACTTTGTTTAGGATGAGGA
GTCTTTGTGTCCCTGTACAGTAGTCTGACGTATTTCCCCTTCTGTCCCCTAGTAAGCCCAGTTGCTGTATCTGAACA
GTTTGAGCTCTTTTTGTAATATACTCTAAACCTGTTATTTCTGTGCTAATAAACGAGATGCAGAACCCTTGAAAAAT
GGA

<210> SEQ ID NO 31

<211> Length : 2,546

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 31

>HSENA78_T5

AGTGGGGAGAGATGAGTGTAGATAAAAGGAGTGCAGAAGGCACGAGGAAGCCACAGTGCTCCGGATCCTCCAATCTT
CGCTCCTCCAATCTCCGCTCCTCCACCCAGTTCAGGAACCCGCGACCGCTCGCAGCGCTCTCTTGACCACTATGAGC
CTCCTGTCCAGCCGCGCGGCCCGTGTCCCCGGTCCTTCGAGCTCCTTGTGCGCGCTGTTGGTGCTGCTGCTGCTGCT
GACGCAGCCAGGGCCCATCGCCAGCGCTGGTCCTGCCGCTGCTGTGTTGAGAGAGCTGCGTTGCGTTTGTTTACAGA
CCACGCAGGGAGTTCATCCCAAAATGATCAGTAATCTGCAAGTGTTCGCCATAGGCCCACAGTGCTCCAAGGTGGAA
GTGGTGTAAGTTCTGTGCTGCTGTGTCCGCTGTGACCTTGGCAAGAGAGAAATCCCGCAGCCTGGGTCTTCAACCTT
GGTATCTCATGAGTGTATCTTCTTTTTCTTTCCTTCAGAGCCTCCCTGAAGAACGGGAAGGAAATTTGTCTTGATCC
AGAAGCCCCTTTTCTAAAGAAAGTCATCCAGAAAATTTTGGACGGTGGAAACAAGGAAAACTGATTAAGAGAAATGA
GCACGCATGGAAAAGTTTCCCAGTCTTCAGCAGAGAAGTTTTCTGGAGGTCTCTGAACCCAGGGAAGACAAGAAGGA
AAGATTTTGTTGTTGTTTGTTTATTTGTTTTTCCAGTAGTTAGCTTTCTTCCTGGATTCCTCACTTTGAAGAGTGTG
AGGAAAACCTATGTTTGCCGCTTAAGCTTTCAGCTCAGCTAATGAAGTGTTTAGCATAGTACCTCTGCTATTTGCTG
TTATTTTATCTGCTATGCTATTGAAGTTTTGGCAATTGACTATAGTGTGAGCCAGGAATCACTGGCTGTTAATCTTT
CAAAGTGTCTTGAATTGTAGGTGACTATTATATTTCCAAGAAATATTCCTTAAGATATTAACTGAGAAGGCTGTGGA

TTTAATGTGGAAATGATGTTTCATAAGAATTCTGTTGATGGAAATACACTGTTATCTTCACTTTTATAAGAAATAGG
AAATATTTTAATGTTTCTTGGGGAATATGTTAGAGAATTTCCTTACTCTTGATTGTGGGATACTATTTAATTATTTC
ACTTTAGAAAGCTGAGTGTTTCACACCTTATCTATGTAGAATATATTTCCTTATTCAGAATTTCTAAAAGTTTAAGT
TCTATGAGGGCTAATATCTTATCTTCCTATAATTTTAGACATTCTTTATCTTTTTAGTATGGCAAACTGCCATCATT
TACTTTTAAACTTTGATTTTATATGCTATTTATTAAGTATTTTATTAGGAGTACCATAATTCTGGTAGCTAAATATA
TATTTTAGATAGATGAAGAAGCTAGAAAACAGGCAAATTCCTGACTGCTAGTTTATATAGAAATGTATTCTTTTAGT
TTTTAAAGTAAAGGCAAACTTAACAATGACTTGTACTCTGAAAGTTTTGGAAACGTATTCAAACAATTTGAATATAA
ATTTATCATTTAGTTATAAAAATATATAGCGACATCCTCGAGGCCCTAGCATTTCTCCTTGGATAGGGGACCAGAGA
GAGCTTGGAATGTTAAAAACAAAACAAAACAAAAAAAAACAAGGAGAAGTTGTCCAAGGGATGTCAATTTTTTATCC
CTCTGTATGGGTTAGATTTTCCAAAATCATAATTTGAAGAAGGCCAGCATTTATGGTAGAATATATAATTATATATA
AGGTGGCCACGCTGGGGCAAGTTCCCTCCCCACTCACAGCTTTGGCCCCTTTCACAGAGTAGAACCTGGGTTAGAGG
ATTGCAGAAGACGAGCGGCAGCGGGGAGGGCAGGGAAGATGCCTGTCGGGTTTTTAGCACAGTTCATTTCACTGGGA
TTTTGAAGCATTTCTGTCTGAATGTAAAGCCTGTTCTAGTCCTGGTGGGACACACTGGGGTTGGGGGTGGGGGAAGA
TGCGGTAATGAAACCGGTTAGTCAGTGTTGTCTTAATATCCTTGATAATGCTGTAAAGTTTATTTTTACAAATATTT
CTGTTTAAGCTATTTCACCTTTGTTTGGAAATCCTTCCCTTTTAAAGAGAAAATGTGACACTTGTGAAAAGGCTTGT
AGGAAAGCTCCTCCCTTTTTTTCTTTAAACCTTTAAATGACAAACCTAGGTAATTAATGGTTGTGAATTTCTATTTT
TGCTTTGTTTTTAATGAACATTTGTCTTTCAGAATAGGATTCTGTGATAATATTTAAATGGCAAAAACAAAACATAA
TTTTGTGCAATTAACAAAGCTACTGCAAGAAAAATAAAACATTTCTTGGTAAAAACGTATGTATTTATATATTATAT
ATTTATATATAATATATATTATATATTTAGCATTGCTGAGCTTTTTAGATGCCTATTGTGTATCTTTTAAAGGTTTT
GACCATTTTGTTATGAGTAATTACATATATATTACATTCACTATATTAAAATTGTACTTTTTTACTATGTGTCTCAT
TGGTT

<210> SEQ ID NO 32

<211> Length : 1,893

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 32

>HUMODCA_T17

GACGTCGGCCCGCCGGCGCCCCACCAGCTCCGCGCGGGCCCGGGTTGGCCACCGCCGGGCCCCCGCCCCTCCCCCGG
CGGTGTCCCGGCCGGAACCGATCGTGGCTGGTTTGAGCTGGTGCGTCTCCATGGCGACCCGCCGGTGCTATAAGTAG
GGAGCGGCGTGCCGTGGGGCTTTGTCAGTCCCTCCTGTAGCCGCCGCCGCCGCCGCCCGCCGCCCCTCTGCCAGCAG
CTCCGGCGCCACCTCGGGCCGGCGTCTCCGGCGGGCGGGAGCCAGGCGCTGACGGGCGCGGCGGGGGCGGCCGAGCG
CTCCTGCGGCTGCGACTCAGGCTCCGGCGTCTGCGCTTCCCCATGGGGCTGGCCTGCGGCGCCTGGGCGCTCTGAGA
TTGTCACTGCTGTTCCAAGGGCACACGCAGAGGGATTTGGAATTCCTGGAGAGTTGCCTTTGTGAGAAGCTGGAAAT

ATTTCTTTCAATTCCATCTCTTAGTTTTCCATAGGAACATCAAGAAATCATGAACAACTTTGGTAATGAAGAGTTTG
ACTGCCACTTCCTCGATGAAGGTTTTACTGCCAAGGACATTCTGGACCAGAAAATTAATGAAGTTTCTTCTTCTGTT
GGTTTTGCGGATTGCCACTGATGATTCCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCACGCTCAGAACCA
GCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATATCGATGTTGTTGGTGTCAGCTTCCATGTAGGAAGCGGCTGT
ACCGATCCTGAGACCTTCGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTTGACATGGGGGCTGAGGTTGGTTTCAG
CATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATTTGAAGAGATCACCGGCG
TAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGACTCTGGAGTGAGAATCATAGCTGAGCCCGGCAGATACTAT
GTTGCATCAGCTTTCACGCTTGCAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGGGCTCTGATGA
CGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTGCATACTCT
ATGACCACGCACATGTAAAGCCCCTTCTGCAAAAGAGACCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGG
GGACCAACATGTGATGGCCTCGATCGGATTGTTGAGCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCT
CTTTGAAAACATGGGCGCTTACACTGTTGCTGCTGCCTCTACGTTCAATGGCTTCCAGAGGCCGACGATCTACTATG
TGATGTCAGGGCCTGCGTGGCAACTCATGCAGCAATTCCAGAACCCCGACTTCCCACCCGAAGTAGAGGAACAGGAT
GCCAGCACCCTGCCTGTGTCTTGTGCCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAGTAT
TAATGTGTAGATAGCACTCTGGTAGCTGTTAACTGCAAGTTTAGCTTGAATTAAGGGATTTGGGGGGACCATGTAAC
TTAATTACTGCTAGTTTTGAAATGTCTTTGTAAGAGTAGGGTCGCCATGATGCAGCCATATGGAAGACTAGGATATG
GGTCACACTTATCTGTGTTCCTATGGAAACTATTTGAATATTTGTTTTATATGGATTTTTATTCACTCTTCAGACAC
GCTACTCAAGAGTGCCCCTCAGCTGCTGAACAAGCATTTGTAGCTTGTACAATGGCAGAATGGGCCAAAAGCTTAGT
GTTGTGACCTGTTTTTAAAATAAAGTATCTTGAAATAATTAGGCA

<210> SEQ ID NO 33

<211> Length : 1,069

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 33

>R00299_T2

GCGGCCGCAGAGCACTTTGCCCGGAGCCCAGCGTCCTCCCCTGAGTTCGCTGAGTCTCCCGGGACCAGCAAAGGCTG
CGCGCCCCGCATCGGCCCGGAGGCGGGGAGCCCTGGGAGGCCTGGCCGAGCTGCCCGCAGGGAAATGGCGGAGAAAG
CGCTTCTCTGCCCGAGTTCAGCCGGGCTGGGGACTTGGCCCTGGGTCCTGAACTCGGCATGGCCAGTTCTGCCTCTG
GCTGTGGACCAGGGTGTGGACTGGAGACCGCGGGGGCCAGTCTCATCGGATCAGATCGAGCAGCTCCATCGGAGATT
TAAGCAGCTGAGTGGAGATCAGCCTACCATTCGCAAGGAGAACTTCAACAATGTCCCGGACCTGGAGCTCAACCCCA
TCCGATCCAAAATTGTTCGTGCCTTCTTCGACAACAGGAACCTGCGCAAGGGACCCAGTGGCCTGGCTGATGAGATC
AATTTCGAGGACTTCCTGACCATCATGTCCTACTTCCGGCCCATCGACACCACCATGGACGAGGAACAGGTGGAGCT
GTCCCGGAAGGAGAAGCTGAGATTTCTGTTCCACATGTACGACTCGGACAGCGACGGCCGCATCACTCTGGAAGAAT

ATCGAAATGTGGTCGAGGAGCTGCTGTCGGGAAACCCTCACATCGAGAAGGAGTCCGCTCGCTCCATCGCCGACGGG
GCCATGATGGAGGCGGCCAGCGTGTGCATGGGGCAGATGGAGCCTGATCAGGTGTACGAGGGGATCACCTTCGAGGA
CTTCCTGAAGATCTGGCAGGGGATCGACATTGAGACCAAGATGCACGTCCGCTTCCTTAACATGGAAACCATGGCCC
TCTGCCACTGACCCACCGCCACCTCCGCGGAGAAACTGCACTTTGCAATGGGGCCGCCTCCCCGCGTAGCTGGAGCA
GCCCAGGCCCGGCGGACAGCCTCTTCCTGCAGCGCCGGTACATAGCCAAGGCTCGTCTGCGCACCTTGTGTCTTGTA
GGGTATGGTATGTGGGACTTCGCTGTTTTTATCTCCAATAAAAAAAAAAAAAAGGTTTGTTAATTAAT

<210> SEQ ID NO 34

<211> Length : 1,250

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 34

>W60282_PEA_1_T11

GGAGCGGCCTAGGGGAGGCCAGGGGCCCACCTGGGCTGGGGCTGTGGAGAGGGAGTGGCTGGGACGGGAGGAAAAAG
AGAGACGGAGATTAGATGGAAGAAGAGGGATTTCAAGACAAATTGCCAGAGATGCAGTCAGAGAGACTGACTGAGAG
ACACAAAGATAGAAGGAATTAGAGAAAGGGCCACACAGAGCCAGACAGAGAGAGAAGAGTGGAGATGGAGACAGGGA
CGAGGACAGAGAAAGGCAGACAGACACATAGGGACAGAAAGAGAAAAATCACACAAAGTCAGAATTACTGAATGACA
GGGAATGACACATAGAACGAGACACAGATTCAGAGACTCAGGGCAGGGAAAGGAAGGCTGCAGACAGACAGACAGAC
AGAGGGAGGCTGAGACACAGGGAGAAGAGGGGCTTGGAGAGGTGGCACAGGCAGGCAGCCAGTGCCTCAGAGGCCTC
CGGGGAGGGCCCTCACACACACCCCGCCCCGGGGCATTAAGGCAGGGCTTGGAGGCCAGTCATCCTGGGCCCGCCCA
GGGCCGCCCCCCTGCCAGCCCGCCTGCCTGGTGCCTGGCACCTGGCGCTCCAACCCAGCCTACCTGCTGTAGCTGCC
GCCACTGCCGTCTCCGCCGCCACTGGGCCCCCAGAGCCCCAGCCCCAGAGCCTAGGAACCTGGGGCCCGCTCCTCCC
CCCTCCAGGCCATGAGGATTCTGCAGTTAATCCTGCTTGCTCTGGCAACAGGGCTTGTAGGGGGAGAGACCAGGATC
ATCAAGGGGTTCGAGTGCAAGCCTCACTCCCAGCCCTGGCAGGCAGCCCTGTTCGAGAAGACGCGGCTACTCTGTGG
GGCGACGCTCATCGCCCCCAGATGGCTCCTGACAGCAGCCCACTGCCTCAAGCCTACGCCTGCCTCACACCTTGCGA
TGCGCCAACATCACCATCATTGAGCACCAGAAGTGTGAGAACGCCTACCCCGGCAACATCACAGACACCATGGTGTG
TGCCAGCGTGCAGGAAGGGGGCAAGGACTCCTGCCAGGGTGACTCCGGGGGCCCTCTGGTCTGTAACCAGTCTCTTC
AAGGCATTATCTCCTGGGGCCAGGATCCGTGTGCGATCACCCGAAAGCCTGGTGTCTACACGAAAGTCTGCAAATAT
GTGGACTGGATCCAGGAGACGATGAAGAACAATTAGACTGGACCCACCCACCACAGCCCATCACCCTCCATTTCCAC
TTGGTGTTTGGTTCCTGTTCACTCTGTTAATAAGAAACCCTAAGCCAAGACCCTCTACGAACATTCTTTGGGCCTCC
TGGACTACAGGAGATGCTGTCACTTAATAATCAACCTGGGGTTCGAAATCAGTGAGACCTGGATTCAAATTCTGCCT
TGAAATATTGTGACTCTGGGAATGACAACACCTGGTTTGTTCTCTGTTGTATCCCCAGCCCCAAAGACAGCTCCTGG
CCATATATCAAGGTTTCAATAAATATTTGCTAAATGAGTGAATC

<210> SEQ ID NO 35

<211> Length : 4,901

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 35

>Z41644_PEA_1_T5

CCTTGCTGTTCATGGCCCAGCAGGGGCCCTATGGGGGTCAGGCCTGCAGGCACTCACACTTGGCACCTGCTCCAAAA
CCCTTTCAGGTCTTTGAGGATCTGAGCCCTGGGCCTGGGTCTCCCGCCGGCTCGGAAAAGCTGGCCTGCCGGGCCAG
ACGAGAGAACCACACGATTCAGAAAAGCAGTGCCCTTCAGCAGCCTCTCCACCGTCTGGGCTCCCCAAAGGCAGAGC
GGGACGCTGGAAATGTGTGCGCGCTGTGGTATGGGTGTGCAAGTGTGCGAAGGCGGCGTGTTGTGTGAGCGAGAGGG
TAGCGGATGTGTGTGTGCGTGTGCGCGCGTGGCTCCGGGTGTGCGCCGCTGCGATAGCGGGTCCTTTCCCGGGGCGG
GCGACGGGCGGGCTGGGAAGGTCTCCTCCCCTCACCACATTGAGAAATCTCAGTGAGTCACCGAGTGGTTCTGCATA
TTAATGAGCTCGCTCGCTGCGAGGGCAGGAGCGGATTTAAAAGAGGCCAGGGCGGGCGGAGGGAGGCTGTGGAGAGA
GCGCGGAGACAAGCGCAGAGCGCAGCGCACGGCCACAGACAGCCCTGGGCATCCACCGACGGCGCAGCCGGAGCCAG
CAGAGCCGGAAGGCGCGCCCCGGGCAGAGAAAGCCGAGCAGAGCTGGGTGGCGTCTCCGGGCCGCCGCTCCGACGGG
CCAGCGCCCTCCCCATGTCCCTGCTCCCACGCCGCGCCCCTCCGGTCAGCATGAGGCTCCTGGCGGCCGCGCTGCTC
CTGCTGCTGCTGGCGCTGTACACCGCGCGTGTGGACGGGTCCAAATGCAAGTGCTCCCGGAAGGGACCCAAGATCCG
CTACAGCGACGTGAAGAAGCTGGAAATGAAGCCAAAGTACCCGCACTGCGAGGAGAAGATGGTTATCATCACCACCA
AGAGCGTGTCCAGGTACCGAGGTCAGGAGCACTGCCTGCACCCCAAGCTGCAGAGCACCAAGCGCTTCATCAAGTGG
TACAACGCCTGGAACGAGAAGCGCAGGTACGCCCCGCCTCTACTCACCTTCCTTCCCACCAGACCCAGCTGTGGCTC
TCAGGATGGGAAGGGACCCCCCCACCAGGTCATCTAGCCCCATCTAATATGTGAACACCCACCACAACATCCACAGC
AAACAGATACTCAGACAATGCTTACATACCCCCAGGGACAAGGAACTCACCACTTACCAAAGCTAGCTATCCATCTC
TTGTCCATTTGCAAGCATGGCAGGTTTGTCATTTTGTAAACTAAAGTCTGTCTCACTCTAATATTTGCATTATAATC
TTAATTCCTCTTTTTATTTCAGTTACGTAAGTTGTTAAATGGCAGAGTGAGCACTGGCATGGCTGCCAGGGGAGCTC
TGAGGACTTCAGTGGGGTGAAATGTGACCACTTAGGTGACTGTGTATGTTGGCTATAAAACTGCGCTATAAAACCAT
GAGGTGCTGAGGATGATCCTTGCCAGAAACATGTTTTCTTCTCCAAGGTGCCCCACTCCCTCTGCTGCCCAGAAACC
TGATAAACTCCTTCCTTCGCAGGTGCTGGAAGGCACCACAGGTTTGGCTCTTTAAAATCAGAGCCACTGTTAACCAA
GGCGGGCAGCAGTGTTAAGACCACCAGCACCCTGAACCAGCCCTGTACTTACTGGGCACTGTTTCCTTAAAATCAGA
AGGTGGCTTCCCATCTCTGGTTTCCTGGGGTCTTATGTCTGTCCTCGGAGGGAGAATCCAGTTCCTAGCTCCCCTGT
ACCATGCGAAGGTAGCCTGTCCTGTCTCACTCCTCAGATACGCAGAGTCTGTTTACACATTTGCCTGCATAGCATGA
TCAGGAAGCACACACACACACACACACACACACACACACGCATGCATGCACACACCATGCAGGTGACTTCCCCAGGA
ACTAGTGCCAGCACCCCTGCTGCAGAGGGGGATATCAAGGCTAAATGGAAGAGAGGGGTGACTTGCCTGGGAGCACA
GGGCAAAGCCAGGACAGCAAACCAGGCCTCCTGGTGCTACCCCACCAGCTGCCCTCACAGGGTGGAAGGTACAGCCA

TAGTGGGTGCCTGCATTGCCCTCCCCTCACCTGGCCCAGCCATGCTACCCCAAGCTCAGCCCTGTGACCAGCTCTCC
CAGAGCTGACACTCGGGCTCAACCCCTATACCTGAGCCTTTTTTGCTGCCTCCAAAACAGCCTCATCTGCAGTTGCT
TGAAATAGAAAGTGATGAGAGCAATAAATTATTTTCTATAAATCTGCTGGGAATGAAGCCCTCTTTCTGGTCAAGCC
AGGCAGCTCATGTGGCAAAGGCCAGAACTGCGCAGTCCACACTCTGTCACCCTCCAGGCCCTGTGTCATCAAAACTG
GCAGAAGGCTAATCCCATGGGCAGGTTATGGAAGAGGCTGAGGGCATCTTGATCTGATTGCTGGGGGATACTCAAAC
CTTTAGCTCACCTTGCTTCTCCCCTCCACCTGAGCTGCAGCCTGGAAAAGGAGGCACCCACAGGTCTAAACATGGCC
CTGCTTTTTTTTTTCCTGAAAATTCCAATAACAAAAGCAACGAGAGCCTCTCACTACCAGGCCTTCTCTCACTTTGC
TATAAAATTAGTTCACCCCTCTTTCTTAGAGTGTTGAGGTCCCTGCCCTCCCCACCTCCCTCCCCTGAAACAAGTTG
AAAATATCTTAATGAACATAGAACAGTGATAAAGGAAGTGTTTGAAGTCCTCTTTGTACAGAGAGAGAGAGAAAGAG
AATGCCAAAGCTAGGTTGGAGGAAGTAGAAGGGTATACGGTGGGCTCAGGCCCATGGGGGCCACACAGAGGAGCTCT
GTGCACTTCAGAGACCAGAGCTTCCAGGGAGCTTCTGGCCACCACAGGAAGCAGCCTAGTCAGGCATTTTATTTCAA
TGGATAATTCAGTGGTCTTACTCAGAAATCAAGAACGAGACAGAAAAGTGATAGGCTAAGTGTAACGTATGGCCCCA
GGGCAGCCATGGGGCAGAACTAGAAGAAAGCAAAATATCTAACTGGGCACAGCTTGAGAGGTGAGGGGAAGGTGGGG
CTGGGAACGAGTAGAGATGAGGCAATGCAGCCAGGAGCAGGGACTGAGGGGCACAGGCCTCCTGCACCACTGCCCCA
CCCCACCAACCACCTCTTCTGTCTCCAGGAAGCAGCTTCTAGAGCTAGCATTCTTCTGGAGGACATGCATTATTTGG
GCAAAATACAAAGAAATATACAAGCCTAAGTCAAGTAAGGGAATGCCTCCCACCCTTGCTATTTTCTCTAAATAGAG
AGGCTGAGTACAGACGCGGAAAGAAACAAGGAGGTGTGGGAGCAGCCCGCCATGCTAGAGAAAGACTACATTCCTGC
CACTAACAGTCGGTGGCCACTGGGCAAATCTTAAGTCTGTGGTGCCTCAGTTTCCTCATATGCAAAGCGGGTTTGTT
CCATAGGCCTCTGAGGACAAAATGAGATTGCAGAAGTGAGATTGCAGATGGTTAGAAAAGACAAAGCCACACTGGTG
TGAGTTTTCATGGTCCCCGGGACCACATCCTCAGAAGGATCCCTCCCACTTCTCCTGGGGGTTCCTGCAGTTCTGGG
ACAGGGGCATTCCCTGCAGACCAGACGTGAATGAAGCCGCTTAGCCAGCATCTTGTGAACGGCCTGCCTCATGTCCT
GAGCCACTTACACATGTGTTTTTTCTCCCCAGGGTCTACGAAGAATAGGGTGAAAAACCTCAGAAGGGAAAACTCCA
AACCAGTTGGGAGACTTGTGCAAAGGACTTTGCAGATTAAAAAAAAAAAAAAAAAAAAGCCTTTCTTTCTCACAGGC
ATAAGACACAAATTATATATTGTTATGAAGCACTTTTTACCAACGGTCAGTTTTTACATTTTATAGCTGCGTGCGAA
AGGCTTCCAGATGGGAGACCCATCTCTCTTGTGCTCCAGACTTCATCACAGGCTGCTTTTTATCAAAAAGGGGAAAA
CTCATGCCTTTCCTTTTTAAAAAATGCTTTTTTGTATTTGTCCATACGTCACTATACATCTGAGCTTTATAAGCGCC
CGGGAGGAACAATGAGCTTGGTGGACACATTTCATTGCAGTGTTGCTCCATTCCTAGCTTGGGAAGCTTCCGCTTAG
AGGTCCTGGCGCCTCGGCACAGCTGCCACGGGCTCTCCTGGGCTTATGGCCGGTCACAGCCTCAGTGTGACTCCACA
GTGGCCCCTGTAGCCGGGCAAGCAGGAGCAGGTCTCTCTGCATCTGTTCTCTGAGGAACTCAAGTTTGGTTGCCAGA
AAAATGTGCTTCATTCCCCCCTGGTTAATTTTTACACACCCTAGGAAACATTTCCAAGATCCTGTGATGGCGAGACA
AATGATCCTTAAAGAAGGTGTGGGGTCTTTCCCAACCTGAGGATTTCTGAAAGGTTCACAGGTTCAATATTTAATGC
TTCAGAAGCATGTGAGGTTCCCAACACTGTCAGCAAAAACCTTAGGAGAAAACTTAAAAATATATGAATACATGCGC
AATACACAGCTACAGACACACATTCTGTTGACAAGGGAAAACCTTCAAAGCATGTTTCTTTCCCTCACCACAACAGA
ACATGCAGTACTAAAGCAATATATTTGTGATTCCCCATGTAATTCTTCAATGTTAAACAGTGCAGTCCTCTTTCGAA
AGCTAAGATGACCATGCGCCCTTTCCTCTGTACATATACCCTTAAGAACGCCCCCTCCACACACTGCCCCCCAGTAT
ATGCCGCATTGTACTGCTGTGTTATATGCTATGTACATGTCAGAAACCATTAGCATTGCATGCAGGTTTCATATTCT
TTCTAAGATGGAAAGTAATAAAATATATTTGAAATGTACCAAAATTCTAGA

<210> SEQ ID NO 36

<211> Length : 3,429

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 36

>Z44808_PEA_1_T11

CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAA
GACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTC
CCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGT
CCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGC
AATGACGACGGCACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAG
GCCCATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAACGCG
AAGGCACAGGAAAAACAGATATTGCATCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAA
ACCAATAAGAATTCAGTGTCATCCTGTGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGA
CAATGTGGTGATCCCTGAGTGTGCGCACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCT
GGTGCGTCCTGGTGGACACGGGGCGCCCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAATGTGACAACACG
GCCAGGGCCCACCCAGCCAAAGCCCGGGACCTGTACAAGGGCCGCCAGCTACAAGGTTGTCCGGGTGCCAAAAAGCA
TGAGTTTCTGACCAGCGTTCTGGACGCGCTGTCCACGGACATGGTCCACGCCGCCTCCGACCCCTCCTCCTCGTCAG
GCAGGCTCTCAGAACCCGACCCCAGCCATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAA
AACTCCAGTGGAGACATCGGCAAAAAGGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAA
ATGTGTGAAGAAGTTTGTTGAATACTGTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCC
TGGGCGTGGCGAAAGAGGACGGCAAAGCGGACACCAAGAAACGCCACACCCCCAGAGGTCATGCTGAAAGTACGTCT
AATAGACAGCCAAGGAAACAAGGATAAATGGCTCATACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCA
CCAAAGAGCAATTAAGAAAACAAAAACAGAAACACATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGT
AGAAATGAGCTATTTAAACAGACTGTTTTAATCTGTGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAAT
GTATGTGTCCTCTTTTGACCTTGGAAATCTGTATGTGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCG

CCTTCCATATGTTAACAGTAGAGCTCTATGCACTCCGGCTGCAATCGTATGGCTTTCTCTAACCCCTGCAGTCACTT
CCAGATGCCTGTGCTTACAGCATTGTGGAATCATGTTGGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCC
GACCCTACAGGCAGTTAACATGCATGGGCTGGTTTGTTTCTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTTCC
AGAGATCTTGCTCATACAATGAATCACGCAACCACTAAAGCTATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAA
TAATATTTTCAAACTGTCGTTGGTGTGATACTTTGGCTCAAAGGATCTTTGCTTTTCCATTTTAAGCTTCTGTTTTG
AGTTTTGCCCTGGGGCTTGAATGAGTCCCAGAGAGTCGTTCGGATGGTGGGAGGCTGCCTAGGAGGCAGTAAATCCA
GTCACAGTGCCTGGGAGGGGCCCATCCTTCCAAAATGTAAATCCAGTCGCGGTGTGACCGAGCTGGCTAACAGGCTT
GTCTGCCTGGTTTTCCTCCTACACGTGGACATTATTCTCCTGATCCTCCTACCTGGTCCACCCCAGGGCTACCGGAA
GGTAAAATCTTCACCTGAACCAATTATGAGCAGTCTCCTTACTGAAGGTACAGCCGGATACGTGGTGCCCCCGGGGC
TGGTGTTGGCAGCCGGGGGGAGGTGCCTGAGGGTCCCCACGGTTCCTTTCTGCTTTTCTGAATGCATCAAGGGTACG
AGAACTTGCCAATGGGAAATTCATCCGAGTGGCACTGGCAGAGAAGGATAGGAGTGGAATGCCCACACAGTGACCAA
CAGAACTGGTCTGCGTGCATAACCAGCTGCCACCCTCAGGCCTGGGCCCCAGAGCTCAGGGCACCCAGTGTCTTAAG
GAACCATTTGGAGGACAGTCTGAGAGCAGGAACTTCAAGCTGTGATTCTATCTCGGCTCAGACTTTTGGTTGGAAAA
AGATCTTCATGGCCCCAAATCCCCTGAGACATGCCTTGTAGAATGATTTTGTGATGTTGTGATGCTTGTGGAGCATC
GCGTAAGGCTTCTTGCTTATTTAAACTGTGCAAGGTAAAAATCAAGCCTTTGGAGCCACAGAACCAGCTCAAGTACA
TGCCAATGTTGTTTAAGAAACAGTTATGATCCTAAACTTTTTGGATAATCTTTTATATTTCTGACCTTTGAATTTAA
TCATTGTTCTTAGATTAAAATAAAATATGCTATTGAAACTA

<210> SEQ ID NO 37

<211> Length : 5,165

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 37

>Z44808_PEA_1_T4

CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAA
GACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTC
CCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGT

CCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGC
AATGACGACGGCACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAG
GCCCATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAACGCG
AAGGCACAGGAAAAACAGATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAGATATTGCA
TCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTG
TGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCTGAGTGTGCGC
ACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGACACGGGGCGC
CCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCG
GGACCTGTACAAGGGCCGCCAGCTACAAGGTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACG
CGCTGTCCACGGACATGGTCCACGCCGCCTCCGACCCCTCCTCCTCGTCAGGCAGGCTCTCAGAACCCGACCCCAGC
CATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACTCCAGTGGAGACATCGGCAAAAA
GGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGTGTGAAGAAGTTTGTTGAATACT
GTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGGCGTGGCGAAAGAGGACGGCAAA
GCGGACACCAAGAAACGCCACACCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACAGGATGCAATGGTGGTGTC
CTCCAGACCCAAAGCCACAACCCATCGCAAGTCAAGAACACTTTCCAGAAGATAAACATGAGTGGGTTCATGTCTCT
CTCCTTCAAAGCCAGGACAAAATCCCCACTTCTTTGCTGCCGCGAGTCAATTTGTGATTTATTTTGTCTGCACCTGT
TTGATGCCAGGTCGACATTTCCTAAGGCAAGCCCCTGTATTTGTTGTGGATTTAAGTGGAGGCGGCCAGCACACACC
TTGGATGTAATTTAAAACCATTTCCTGAGGAAAGATGTGTGATATGCTTTCCTTTGTTTAGCAAATGTTTATGGTTT
TAACTTTAAATCTCACCGCAAATCACTTACACTTGAAAACAGGGCTGGTCTGAAAGTAATTACCCTCCCTGAGTGCC
AAGACCTCCAGAAGTTGTTTTCATTCCCGAATGGCAATCACTGTACTCATGCGCTCCACGCATCTTAAATAAACTCA
GTTCAAAGCACATGCCTCCTGCTTCAGCTCTTTTTTCCAAAAAGAGAAACAGAAGCAGGTTCCCCCTCCTTTTATAG
TGCCTGCGTGGACACGCGGACCTCCATGCCTTTCATGCTGTGGCTATGTCAGCAAACTACGATATTGGGATGATCCT
AACGGGCAAGCCAGCTGCGGCTCCTACCGGCCGTGGCCATTGAAGGCCACCATGTTGCTTTGAAACATCTCAAAGAA
TAACATAGTGCCAGCCAGCAAGGGTTTCACCATATGCATGACCCAGACAGGAACTATCAAAAGAAGGGATCACGGGA
AGGTGCATGATGCTAATGTGGAATCCAGAGGAGCTCTTTCCTGATCTCTTCAGCTTCCGCTGCCACTCCAGAATCAT
CAGAGCTGATATTAAATAAGTTAAAATGTTAGTCCACCGTCTCCTCCTGCAATCCTAACCATCTTTTGAGACTGTTA
GAATACTTTGACGGGTTGTCTTTCTGTGCAACTAATTTAAACCTCAAGTTTAGTGTAGGAGATGGGTTTGTCTTCTC
ACCTCTTCAGATCTTTATCAAGGGGGAATAAAAGCCAACCCAGAAACCTAAACTTTAAAATTTAATTATTTGAAATA
ATAAAACAGAAGAAGGGATCAACATTTGTCGGAATTGGCACTCTTGGAAAACTAAGTCTAGGAGATCATATATTGCT
TTTTTTTTTTCATTCTAAATTACTTTTAATTGAAAGTCAAGATGCTGAGTTACAGTTGTTTATCATTATAATAAGCA
AACTTTTTAAGTTGGATTTCTTCTTAAAGAGGTAAACTAGTGAACAAAAAAGATAAAAAGGAAAATTAAGAATCAAC
TATGCCTTTATCAAATTTGAAGCATAAGTTATATTATTAAAATTATTTTTGTATAATCAAGGTGATAAGACATTCTG
GAAAACATTTAATGTATTTAGTACTTAGAATATTTACAGTGGATGTTACTTTTTTGAAACGATATATTTTTCCCAAT
TTTTCTATCATGTCAAGGAAGGAAACTGTTAAGAAGTTACCAGTGTCCAAAATGTCTTCATTGTTTCTTACTCATAC
TTACACCTCACATGACCTGCCCAGCCCTCTTTGGTTCAGTTCATTCCCAGAAGCCAAGCCTTAGTCTTCACAGATGA
GCGACACACACCTCTGAATATAATGTCTCTTTTTTGTTTTTTCCTTTTCAGCCAAGGAAACAAGGATAAATGGCTCA
TACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCACCAAAGAGCAATTAAGAAAACAAAAACAGAAACAC
ATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGAGCTATTTAAACAGACTGTTTTAATCTG

TGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAATGTATGTGTCCTCTTTTGACCTTGGAAATCTGTATG
TGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCGCCTTCCATATGTTAACAGTAGAGCTCTATGCACTC
CGGCTGCAATCGTATGGCTTTCTCTAACCCCTGCAGTCACTTCCAGATGCCTGTGCTTACAGCATTGTGGAATCATG
TTGGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCCGACCCTACAGGCAGTTAACATGCATGGGCTGGTTT
GTTTCTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTTCCAGAGATCTTGCTCATACAATGAATCACGCAACCAC
TAAAGCTATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAATAATATTTTCAAACTGTCGTTGGTGTGATACTTTG
GCTCAAAGGATCTTTGCTTTTCCATTTTAAGCTTCTGTTTTGAGTTTTGCCCTGGGGCTTGAATGAGTCCCAGAGAG
TCGTTCGGATGGTGGGAGGCTGCCTAGGAGGCAGTAAATCCAGTCACAGTGCCTGGGAGGGGCCCATCCTTCCAAAA
TGTAAATCCAGTCGCGGTGTGACCGAGCTGGCTAACAGGCTTGTCTGCCTGGTTTTCCTCCTACACGTGGACATTAT
TCTCCTGATCCTCCTACCTGGTCCACCCCAGGGCTACCGGAAGGTAAAATCTTCACCTGAACCAATTATGAGCAGTC
TCCTTACTGAAGGTACAGCCGGATACGTGGTGCCCCCGGGGCTGGTGTTGGCAGCCGGGGGGAGGTGCCTGAGGGTC
CCCACGGTTCCTTTCTGCTTTTCTGAATGCATCAAGGGTACGAGAACTTGCCAATGGGAAATTCATCCGAGTGGCAC
TGGCAGAGAAGGATAGGAGTGGAATGCCCACACAGTGACCAACAGAACTGGTCTGCGTGCATAACCAGCTGCCACCC
TCAGGCCTGGGCCCCAGAGCTCAGGGCACCCAGTGTCTTAAGGAACCATTTGGAGGACAGTCTGAGAGCAGGAACTT
CAAGCTGTGATTCTATCTCGGCTCAGACTTTTGGTTGGAAAAAGATCTTCATGGCCCCAAATCCCCTGAGACATGCC
TTGTAGAATGATTTTGTGATGTTGTGATGCTTGTGGAGCATCGCGTAAGGCTTCTTGCTTATTTAAACTGTGCAAGG
TAAAAATCAAGCCTTTGGAGCCACAGAACCAGCTCAAGTACATGCCAATGTTGTTTAAGAAACAGTTATGATCCTAA
ACTTTTTGGATAATCTTTTATATTTCTGACCTTTGAATTTAATCATTGTTCTTAGATTAAAATAAAATATGCTATTG
AAACTA

<210> SEQ ID NO 38

<211> Length : 3,575

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 38

>Z44808_PEA_1_T5

CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAA

GACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTC
CCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGT
CCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGC
AATGACGACGGCACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAG
GCCCATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAACGCG
AAGGCACAGGAAAAACAGATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAGATATTGCA
TCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTG
TGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCTGAGTGTGCGC
ACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGACACGGGGCGC
CCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCG
GGACCTGTACAAGGGCCGCCAGCTACAAGGTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACG
CGCTGTCCACGGACATGGTCCACGCCGCCTCCGACCCCTCCTCCTCGTCAGGCAGGCTCTCAGAACCCGACCCCAGC
CATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACTCCAGTGGAGACATCGGCAAAAA
GGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGTGTGAAGAAGTTTGTTGAATACT
GTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGGCGTGGCGAAAGAGGACGGCAAA
GCGGACACCAAGAAACGCCACAGAAGTAAGAGAAACCTGTGATGGCCAGAGCCCAGATGTTCTTAGGAGGCAAGCCA
GGAGAAGCCGGGTCTGACTTTTCAGCTCAGAGACAGCACTCCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACA
GCCAAGGAAACAAGGATAAATGGCTCATACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCACCAAAGAG
CAATTAAGAAAACAAAAACAGAAACACATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGA
GCTATTTAAACAGACTGTTTTAATCTGTGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAATGTATGTGT
CCTCTTTTGACCTTGGAAATCTGTATGTGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCGCCTTCCAT
ATGTTAACAGTAGAGCTCTATGCACTCCGGCTGCAATCGTATGGCTTTCTCTAACCCCTGCAGTCACTTCCAGATGC
CTGTGCTTACAGCATTGTGGAATCATGTTGGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCCGACCCTAC
AGGCAGTTAACATGCATGGGCTGGTTTGTTTCTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTTCCAGAGATCT
TGCTCATACAATGAATCACGCAACCACTAAAGCTATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAATAATATTT
TCAAACTGTCGTTGGTGTGATACTTTGGCTCAAAGGATCTTTGCTTTTCCATTTTAAGCTTCTGTTTTGAGTTTTGC
CCTGGGGCTTGAATGAGTCCCAGAGAGTCGTTCGGATGGTGGGAGGCTGCCTAGGAGGCAGTAAATCCAGTCACAGT
GCCTGGGAGGGGCCCATCCTTCCAAAATGTAAATCCAGTCGCGGTGTGACCGAGCTGGCTAACAGGCTTGTCTGCCT
GGTTTTCCTCCTACACGTGGACATTATTCTCCTGATCCTCCTACCTGGTCCACCCCAGGGCTACCGGAAGGTAAAAT
CTTCACCTGAACCAATTATGAGCAGTCTCCTTACTGAAGGTACAGCCGGATACGTGGTGCCCCCGGGGCTGGTGTTG
GCAGCCGGGGGGAGGTGCCTGAGGGTCCCCACGGTTCCTTTCTGCTTTTCTGAATGCATCAAGGGTACGAGAACTTG
CCAATGGGAAATTCATCCGAGTGGCACTGGCAGAGAAGGATAGGAGTGGAATGCCCACACAGTGACCAACAGAACTG
GTCTGCGTGCATAACCAGCTGCCACCCTCAGGCCTGGGCCCCAGAGCTCAGGGCACCCAGTGTCTTAAGGAACCATT
TGGAGGACAGTCTGAGAGCAGGAACTTCAAGCTGTGATTCTATCTCGGCTCAGACTTTTGGTTGGAAAAAGATCTTC
ATGGCCCCAAATCCCCTGAGACATGCCTTGTAGAATGATTTTGTGATGTTGTGATGCTTGTGGAGCATCGCGTAAGG
CTTCTTGCTTATTTAAACTGTGCAAGGTAAAAATCAAGCCTTTGGAGCCACAGAACCAGCTCAAGTACATGCCAATG
TTGTTTAAGAAACAGTTATGATCCTAAACTTTTTGGATAATCTTTTATATTTCTGACCTTTGAATTTAATCATTGTT
CTTAGATTAAAATAAAATATGCTATTGAAACTA

<210> SEQ ID NO 39

<211> Length : 2,397

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 39

>Z44808_PEA_1_T8

CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAA
GACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTC
CCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGT
CCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGC
AATGACGACGGCACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAG
GCCCATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAACGCG
AAGGCACAGGAAAAACAGATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAGATATTGCA
TCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTG
TGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCTGAGTGTGCGC
ACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGACACGGGGCGC
CCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCG
GGACCTGTACAAGGGCCGCCAGCTACAAGGTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACG
CGCTGTCCACGGACATGGTCCACGCCGCCTCCGACCCCTCCTCCTCGTCAGGCAGGCTCTCAGAACCCGACCCCAGC
CATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACTCCAGTGGAGACATCGGCAAAAA
GGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGTGTGAAGAAGTTTGTTGAATACT
GTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGGCGTGGCGAAAGAGGACGGCAAA
GCGGACACCAAGAAACGCCACACCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACAGGATGCAATGGTGGTGTC
CTCCAGACCCAAAGCCACAACCCATCGCAAGTCAAGAACACTTTCCAGAAGATAAACATGAGTGGGTTCATGTCTCT
CTCCTTCAAAGCCAGGACAAAATCCCCACTTCTTTGCTGCCGCGAGTCAATTTGTGATTTATTTTGTCTGCACCTGT
TTGATGCCAGGTCGACATTTCCTAAGGCAAGCCCCTGTATTTGTTGTGGATTTAAGTGGAGGCGGCCAGCACACACC

TTGGATGTAATTTAAAACCATTTCCTGAGGAAAGATGTGTGATATGCTTTCCTTTGTTTAGCAAATGTTTATGGTTT
TAACTTTAAATCTCACCGCAAATCACTTACACTTGAAAACAGGGCTGGTCTGAAAGTAATTACCCTCCCTGAGTGCC
AAGACCTCCAGAAGTTGTTTTCATTCCCGAATGGCAATCACTGTACTCATGCGCTCCACGCATCTTAAATAAACTCA
GTTCAAAGCA

<210> SEQ ID NO 40

<211> Length : 2,206

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 40

>Z44808_PEA_1_T9

CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAA
GACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTC
CCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGT
CCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGC
AATGACGACGGCACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAG
GCCCATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAACGCG
AAGGCACAGGAAAAACAGATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAGATATTGCA
TCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAGTGTCATCCTG
TGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCTGAGTGTGCGC
ACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGACACGGGGCGC
CCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCG
GGACCTGTACAAGGGCCGCCAGCTACAAGGTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACG
CGCTGTCCACGGACATGGTCCACGCCGCCTCCGACCCCTCCTCCTCGTCAGGCAGGCTCTCAGAACCCGACCCCAGC
CATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACTCCAGTGGAGACATCGGCAAAAA
GGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGTGTGAAGAAGTTTGTTGAATACT
GTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGGCGTGGCGAAAGAGGACGGCAAA

GCGGACACCAAGAAACGCCACACCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACAGCTACTGTGGTTGAGAGG
AAAGGTGTCTTTTTATTGCTTCTAGAGACGTTGAAAGTGTGACCTGAGCACCTCATAATCATGTGAAAAAGACACTC
AAAAACTACCATTTGAATGGATGGATGAAAATAACCTCCGTATATTCTACGAAGATGTTTAATAATAAATAGGTTTC
GTTATAAGAGAATGTGTGTCACTTCGTCTCTTCCCTCACCCCCGAGACTTAGTGACAGTTATTTTTGACTTTTCCAA
CTATACTATTTGCCTAGAAAATGTGTCTATTAAATAGCGTATTGAGAAAT

<210> SEQ ID NO 41

<211> Length : 1,173

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 41

>AA161187_T0

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA
GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT
GGACTCAGGAAGCCGGAGTCGCAGGAGGCGGCGCCCTTATCAGGACCATGCGGCCGACGGGTCATCACGTCGCGCAT
CGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCACGTATGCG
GAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCC
TCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTA
CTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTG
CACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGAC
TGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAAGTTCAGGT
CGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGGAGACATGG
TTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTCGGTGACTCAGGTGGACCCTTGGCCTGTAACAAGAAT
GGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGGCTGTGGTCGGCCCAATCGGCCCGGTGTCTACACCAA
TATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCTCCTGGCCAC
TACTCTTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCATGCAGCCTGGGG
CCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAAACACATTCCAGTTGATGCCTTGCAGGGCA
TTCTTCAAAAGCAATGGC

<210> SEQ ID NO 42

<211> Length : 1,104

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 42

>AA161187_T7

GCACACACGCGAGGGGACCCTGGGTGGGCAAAAACGTGCTTTCCCGGACGGGGTTGAAGGGGAGAAAGGGAGAGGTC
GGGCTTGGGGGGCTGCCTCCCGCGGCTCAGCAGTTCCTCTGACCATCCGAGGACCATGCGGCCGACGGGTCATCACG
TCGCGCATCGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCA
CGTATGCGGAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTA
GTGATCCCTCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTAC
ACCCGTTACTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAA
GCTGTCTGCACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACC
GGACAGACTGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAA
GTTCAGGTCGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGG
AGACATGGTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTCGGTGACTCAGGTGGACCCTTGGCCTGTA
ACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGGCTGTGGTCGGCCCAATCGGCCCGGTGTC
TACACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCTC
CTGGCCACTACTCTTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCATGCA
GCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAAACACATTCCAGTTGATGCCTT
GCAGGGCATTCTTCAAAAGCAATGGC

<210> SEQ ID NO 43

<211> Length : 1,105

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 43

>AA161187_T15

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA
GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT
GGACTCAGGAAGCCGGAGTCGCAGGAGGCGGCGCCCTTATCAGGACCATGCGGCCGACGGGTCATCACGTCGCGCAT
CGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCACGTATGCG
GAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCC
TCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTA
CTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTG
CACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGAC

TGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGGAAGTTCAGGTCGCCATCATAAACAACTCTATGTGCA
ACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGGAGACATGGGTGACTCAGGTGGACCCTTGGCCTGT
AACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGGCTGTGGTCGGCCCAATCGGCCCGGTGT
CTACACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCT
CCTGGCCACTACTCTTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCATGC
AGCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAAACACATTCCAGTTGATGCCT
TGCAGGGCATTCTTCAAAAGCAATGGC

<210> SEQ ID NO 44

<211> Length : 1,466

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 44

>AA161187_T16

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA
GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT
GGACTCAGGAAGCCGGAGTCGCAGGAGGCGGCGCCCTTATCAGGACCATGCGGCCGACGGGTCATCACGTCGCGCAT
CGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCACGTATGCG
GAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCC
TCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTA
CTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTG
CACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGAC
TGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGGTTGCTGTCTCTCTCCTTCCCACTATCGTCCGCACAG
CACTGCCATCTCCCCACACCCTCCAGGAAGTTCAGGTCGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTC
AAGTACAGTTTCCGCAAGGACATCTTTGGAGACATGGTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTT
CGTGAGTGTCCCTGCCACCACTCCCAGCCCAGGAAAGCATCCTGTGTCCCTGTGCCTTATTTGACCCTCATGCCAAC
CCCGGGAGGTGGAGACTGTTGCCCCACTCTGCAGATGCAGAAACGGAGGCTTGGCTGCTGCCAGGGGGAGGAGGAGG
ATGTGCACCCAGTCTACCCAGCCCCATAGCCCTTCCCACTCTCAGCCCCTCCCCTGCCCCACTCACTCTGCCCCAGG
CTGACCTCAGCCCCGCTGCTCCCCAGGGTGACTCAGGTGGACCCTTGGCCTGTAACAAGAATGGACTGTGGTATCAG
ATTGGAGTCGTGAGCTGGGGAGTGGGCTGTGGTCGGCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTT
TGAGTGGATCCAGAAGCTGATGGCCCAGAGTGGCATGTCCCAGCCAGACCCCTCCTGGCCACTACTCTTTTTCCCTC
TTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGAGCCTACCTGAGCCCATGCAGCCTGGGGCCACTGCCAAGTCAG
GCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAAACACATTCCAGTTGATGCCTTGCAGGGCATTCTTCAAAAGCAAT
GGC

<210> SEQ ID NO 45

<211> Length : 1,354

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 45

>AA161187_T20

GCACACACGCGAGGGGACCCTGGGTGGGCAAAAACGTGCTTTCCCGGACGGGGTTGAAGGGGAGAAAGGGAGAGGTC
GGGCTTGGGGGGCTGCCTCCCGCGGCTCAGCAGTTCCTCTGACCATCCGAGGACCATGCGGCCGACGGGTCATCACG
TCGCGCATCGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCA
CGTATGCGGAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACTGACCTTAGTGATC
CCTCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGT
TACTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTC
TGCACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAG
ACTGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGCACTGCCATCTCCCCACACCCTCCAGGAAGTTCAG
GTCGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTCAAGTACAGTTTCCGCAAGGACATCTTTGGAGACAT
GGTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTCGTGAGTGTCCCTGCCACCACTCCCAGCCCAGGAA
AGCATCCTGTGTCCCTGTGCCTTATTTGACCCTCATGCCAACCCCGGGAGGTGGAGACTGTTGCCCCACTCTGCAGA
TGCAGAAACGGAGGCTTGGCTGCTGCCAGGGGGAGGAGGAGGATGTGCACCCAGTCTACCCAGCCCCATAGCCCTTC
CCACTCTCAGCCCCTCCCCTGCCCCACTCACTCTGCCCCAGGCTGACCTCAGCCCCGCTGCTCCCCAGGGTGACTCA
GGTGGACCCTTGGCCTGTAACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGGCTGTGGTCG
GCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATGGCCCAGAGTGGCA
TGTCCCAGCCAGACCCCTCCTGGCCACTACTCTTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGGCCGGTCTGA
GCCTACCTGAGCCCATGCAGCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTGGTAATAAAC
ACATTCCAGTTGATGCCTTGCAGGGCATTCTTCAAAAGCAATGGC

<210> SEQ ID NO 46

<211> Length : 1,171

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 46

>AA161187_T21

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA
GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT
GGACTCAGGAAGCCGGAGTCGCAGGAGGCGGCGCCCTTATCAGGACCATGCGGCCGACGGGTCATCACGTCGCGCAT
CGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCACGTATGCG
GAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCC
TCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTA
CTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTG
CACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGAC
TGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGATAAGAGGACACAGTGAGAAGATGGGGGTCTGCTCGC
CAGGAAGGAGCCCTCACCAGCAGCCGCATCGCTCAGCACCTTGATCCTGGACTTCCAGCCTCCAGAGCTGTGAGAAA
CAAACCTCTATCATCTACCAGCCGCCCACGGCGTGGGATTTGTGTTACAGCAGCCTGAGCTGACCCAGACGCCAAGG
AGCAACACACGCACCAGGGTAGGCTGGAGAAACCAGAACCCGGGAATCCCGCCTCCCTCAACTTGAAACTTGGGAAT
AGTGTATTCTCTTTTCAACACTTGCACTAGTAGAAGGTTAATTACATGAAAGATTAGGCAAAATGTATGGCTATGTG
TCCTGGTTTTCCAATAAAAGTATTGAGTTTCTCTGGGGAAAGTGCAGATAAAATGCTTAGTGGAGGCTGGGCGCTGT
GGCTTATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGCAGGCAGATCACAAGGTCAGGAGTTTGAGACCGG
CCTGGCCAATATGATG

<210> SEQ ID NO 47

<211> Length : 953

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 47

>AA161187_T22

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA
GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT
GGACTCAGGAAGCCGGAGTCGCAGGAGGCGGCGCCCTTATCAGGACCATGCGGCCGACGGGTCATCACGTCGCGCAT
CGTGGGTGGAGAGGACGCCGAACTCGGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCCCACGTATGCG
GAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAACCTATAGTGACCTTAGTGATCCC
TCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAGGCCTACTACACCCGTTA
CTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGCCTTGGTGAAGCTGTCTG
CACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGTTTGAGAACCGGACAGAC
TGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGGTGCTCACCAATGCCCCAGGCATCAGGCTCCTGGGCTG
CCTCTCCATGCCTCCCACACCCACCCTAGCTCTGGCCGATTCTCCTGCAGCACTGAGCCCATTCCTCTCCCCAGAAA
CTTCCAAGCCATGCTCAACCGCAGCTCCCACGGAAACCCCTCTGGGGGTTCCTCTGGTGGGCCTGCCCTGGCACCTG

CGTGTCCCCCAACACACATGCCCTGAAAGAAGTGGGCCCAGCATCCGGAGGAGCCCCGGCAGCCCCAGACTGGGCGT
GTTCCCTGTATCAGGAATCCCTTCCCTCT

<210> SEQ ID NO 48

<211> Length : 3,110

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 48

>R66178_T2

GCAGCGAGCGCGGCTCCACATTGTTGCGGATCGCCGGCACCCGGCAGAGCGGCGGCGGCTGGGACGCGCGGCGCCTC
CGACCCGTTCTCCTCGCGCCCGGCCGCGCAGCCAGAGCCACCCGGGCCGCCGACCGCCGAGCCCCGCGCCCGCCGCC
TGGGCCCCGAGCCTTCTGCGCCGCCCGGGTGCGTCCCGCCACCCTCGGAGGACGGCCGGCCATGGACGCCTGCAAGT
TGGAGCCGAGCGGGAGGGTGTGAGCGGGCCGGGGCCAGGAGCCCGCGCCGCGCAACCGGGCAGCCGGGCGCGCCGGG
GGTGGGTCCCTCTCCCCAGCCCCGCTCTGCGTGGAAGAAGAGGGCGGGGACCGGCGCCGGGAGGAGAGCGGAGGAGG
CGAAGGGGCATGACTCGTGCAACTTGCGGCGGGCATCTGCCGAGCCTCTGAGCCGGCGGCGGCCCGGGGCCCGGACT
GCGGCCGCGCGGATCCACCCAGCCCACCCCGCCCCGGCCGACGGCTGCAGCTGACCTGGATCCTTCGAGCGCCCGCC
GACCGCCAGCGATCTTCCCTCATCTTCCGGGCTGGTTTCTGCTGCGCGAGGAGCGCTGCCCTCGCCGCCCCTCTCGC
CGGACCCCCGGCCCCCGATGGCTCGGATGGGGCTTGCGGGCGCCGCTGGACGCTGGTGGGGACTCGCTCTCGGCTTG
ACCGCATTCTTCCTCCCAGGCGTCCACTCCCAGGTGGTCCAGGTGAACGACTCCATGTATGGCTTCATCGGCACAGA
CGTGGTTCTGCACTGCAGCTTTGCCAACCCGCTTCCCAGCGTGAAGATCACCCAGGTCACATGGCAGAAGTCCACCA
ATGGCTCCAAGCAGAACGTGGCCATCTACAACCCATCCATGGGCGTGTCCGTGCTGGCTCCCTACCGCGAGCGTGTG
GAATTCCTGCGGCCCTCCTTCACCGATGGCACTATCCGCCTCTCCCGCCTGGAGCTGGAGGATGAGGGTGTCTACAT
CTGCGAGTTTGCTACCTTCCCTACGGGCAATCGAGAAAGCCAGCTCAATCTCACGGTGATGGCCAAACCCACCAATT
GGATAGAGGGTACCCAGGCAGTGCTTCGAGCCAAGAAGGGGCAGGATGACAAGGTCCTGGTGGCCACCTGCACCTCA
GCCAATGGGAAGCCTCCCAGTGTGGTATCCTGGGAAACTCGGTTAAAAGGTGAGGCAGAGTACCAGGAGATCCGGAA
CCCCAATGGCACAGTGACGGTCATCAGCCGCTACCGCCTGGTGCCCAGCAGGGAAGCCCACCAGCAGTCCTTGGCCT
GCATCGTCAACTACCACATGGACCGCTTCAAGGAAAGCCTCACTCTCAACGTGCAGTATGAGCCTGAGGTAACCATT
GAGGGGTTTGATGGCAACTGGTACCTGCAGCGGATGGACGTGAAGCTCACCTGCAAAGCTGATGCTAACCCCCCAGC
CACTGAGTACCACTGGACCACGCTAAATGGCTCTCTCCCCAAGGGTGTGGAGGCCCAGAACAGAACCCTCTTCTTCA
AGGGACCCATCAACTACAGCCTGGCAGGGACCTACATCTGTGAGGCCACCAACCCCATCGGTACACGCTCAGGCCAG
GTGGAGGTCAATATCACAGGTGAGGGTCACAGTCTGCCCATCAGTCCTGGAGTTCTTCAGACCCAGAATTGCGGGCC
TTGAAGGCCAGTGTCCTGGAAGGGAGGCAGGTGTGAGCCTGCAAGTGTGCATGCCCAGCCATGGTATGGACATGTGT
CTCTGGGCATGTAAATGTGAACCAGTGTGAACAGGCTCCGTCTGCATGCTGAGTGTGCATGTGGGAGCCCGTGGCTG
TGCCGTGGCAACGTGCCATTCTCTGAGCCAGCGAATGGCAGTGTGTTGGGAGGTCTGAGAAGGCAGCTGCATCCGTG

CCTCTGGGAGGATTCGGTTCTCCCCCAGCTTGCCGAGGCCCTGCCTGATGGTCTGACACGAGGCACAGCTGCTGCAG
CTGCAGATGGACAGAAGGGCTTCCCAGAGGTGGACCCCAGGCCTCCCCACTCTCCCTGTGGCTGGCTGCACTGCATG
CTGGGGGGTGTAGTTCTTGCAGCTTCCAGGCCTAATCTGATGCCGGAGCATTTCCTGCCTGAGGAAGCGCCAGGCAT
TGGTTTCGGAGGCAAACCCAAACATTCTCTTTGACCCCAAACCTCCAGATCCTAGATCCAGACTGTAAGCCCTAACA
CTTCACTCCACCTCAGATCTATCCAAAGCCCCCAGCACCAGCCCACCCACCTCAGTCAGAGAACCAGGACCCCAAAG
GCATGCAGAGCCCCCACTTCCCCACTGTCTTGGCCAGCCAGGGACCCCAGAAGAGAGGTTACAACCCTTCAGGAATA
GGGACAAGCTGCTCCCTTTGTAAGAGGATGTGAGGGAGGCTGGCTGGGCCCCTGCCAGCAAACACAAATGACCCTGC
GGCCTGGCTCTTCTCTCTCCTCCCAGCTGCGGCCCTTGCAGCTCTGCTCCTGGCACAGAGACAGGAGCTACTGGCTG
AGTGTAACAGCTGGAGGGATGGAGGGGGAGGGGAGGACGCTCCACTCCACGCCAGACAGCCCCTTCTGCTTGCAAAT
GAGTTAGATCCCCATGCTTCTCCTTTCTTCTCTCCCTCACTCAGTATCCTCACTCGAAAGTCTTTGATGCTGGAAGG
TCATCCCAGCCATTCTGCTGCTGCTACACAGGCCCAGCCCTAAACAAAATAACCGGGGTTCTTTGGTCCCAAAAGAT
CCCCAGGAAAGAGTAAACCTCCTTAAGACTTAAGGAAAAAAGTTGGCTAGGCGTGGTGGCTCACACCTGTAATCCCA
GCACTTTGGGAGGCCGAGGTGGGCAGACCACTTGAGATCAGGAGTTCGAGACCAGCCTGGCCAACGTTGTGAAACCC
CGTCTCTACAAAATATTAGCCGGGTGTGGCAGTGCGTGCCTGTAGTCCCAACTACTCAGGAGGCTGAGGCACAAGAA
TTGCTTGAGCCTGGGAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCCAGGCGATGGAGCGAG
ACTCTGTCCCGCAAAAAAAAAAAAAAAAA

<210> SEQ ID NO 49

<211> Length : 1,903

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 49

>R66178_T3

GCAGCGAGCGCGGCTCCACATTGTTGCGGATCGCCGGCACCCGGCAGAGCGGCGGCGGCTGGGACGCGCGGCGCCTC
CGACCCGTTCTCCTCGCGCCCGGCCGCGCAGCCAGAGCCACCCGGGCCGCCGACCGCCGAGCCCCGCGCCCGCCGCC
TGGGCCCCGAGCCTTCTGCGCCGCCCGGGTGCGTCCCGCCACCCTCGGAGGACGGCCGGCCATGGACGCCTGCAAGT
TGGAGCCGAGCGGGAGGGTGTGAGCGGGCCGGGGCCAGGAGCCCGCGCCGCGCAACCGGGCAGCCGGGCGCGCCGGG
GGTGGGTCCCTCTCCCCAGCCCCGCTCTGCGTGGAAGAAGAGGGCGGGGACCGGCGCCGGGAGGAGAGCGGAGGAGG
CGAAGGGGCATGACTCGTGCAACTTGCGGCGGGCATCTGCCGAGCCTCTGAGCCGGCGGCGGCCCGGGGCCCGGACT
GCGGCCGCGCGGATCCACCCAGCCCACCCCGCCCCGGCCGACGGCTGCAGCTGACCTGGATCCTTCGAGCGCCCGCC
GACCGCCAGCGATCTTCCCTCATCTTCCGGGCTGGTTTCTGCTGCGCGAGGAGCGCTGCCCTCGCCGCCCCTCTCGC
CGGACCCCCGGCCCCCGATGGCTCGGATGGGGCTTGCGGGCGCCGCTGGACGCTGGTGGGGACTCGCTCTCGGCTTG
ACCGCATTCTTCCTCCCAGGCGTCCACTCCCAGGTGGTCCAGGTGAACGACTCCATGTATGGCTTCATCGGCACAGA
CGTGGTTCTGCACTGCAGCTTTGCCAACCCGCTTCCCAGCGTGAAGATCACCCAGGTCACATGGCAGAAGTCCACCA
ATGGCTCCAAGCAGAACGTGGCCATCTACAACCCATCCATGGGCGTGTCCGTGCTGGCTCCCTACCGCGAGCGTGTG

GAATTCCTGCGGCCCTCCTTCACCGATGGCACTATCCGCCTCTCCCGCCTGGAGCTGGAGGATGAGGGTGTCTACAT
CTGCGAGTTTGCTACCTTCCCTACGGGCAATCGAGAAAGCCAGCTCAATCTCACGGTGATGGCCAAACCCACCAATT
GGATAGAGGGTACCCAGGCAGTGCTTCGAGCCAAGAAGGGGCAGGATGACAAGGTCCTGGTGGCCACCTGCACCTCA
GCCAATGGGAAGCCTCCCAGTGTGGTATCCTGGGAAACTCGGTTAAAAGGTGAGGCAGAGTACCAGGAGATCCGGAA
CCCCAATGGCACAGTGACGGTCATCAGCCGCTACCGCCTGGTGCCCAGCAGGGAAGCCCACCAGCAGTCCTTGGCCT
GCATCGTCAACTACCACATGGACCGCTTCAAGGAAAGCCTCACTCTCAACGTGCAGTATGAGCCTGAGGTAACCATT
GAGGGGTTTGATGGCAACTGGTACCTGCAGCGGATGGACGTGAAGCTCACCTGCAAAGCTGATGCTAACCCCCCAGC
CACTGAGTACCACTGGACCACGCTAAATGGCTCTCTCCCCAAGGGTGTGGAGGCCCAGAACAGAACCCTCTTCTTCA
AGGGACCCATCAACTACAGCCTGGCAGGGACCTACATCTGTGAGGCCACCAACCCCATCGGTACACGCTCAGGCCAG
GTGGAGGTCAATATCACAGCTTTCTGTCAACTTATCTATCCGGGCAAAGGGAGGACAAGAGCTAGGATGTTCTGAGG
AGAGACTTCACCTGGGACGTGAAAGGAGCATGGGCTTGATGTCAGACAGCTGTGACCCTGGACAGGGCCCCCCCCAC
CATCTGTAAAACGGGGACAGTATGATGTACCTTGAAGGGCTGTTGTCAGAATTCTACGTGATGTAAGTCAAGCACCT
AGCACAGATCAGTCTGTCAATAAATGGCCAATGTTCCGTGATTATTACCCTGACC

<210> SEQ ID NO 50

<211> Length : 2,364

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 50

>R66178_T7

GCAGCGAGCGCGGCTCCACATTGTTGCGGATCGCCGGCACCCGGCAGAGCGGCGGCGGCTGGGACGCGCGGCGCCTC
CGACCCGTTCTCCTCGCGCCCGGCCGCGCAGCCAGAGCCACCCGGGCCGCCGACCGCCGAGCCCCGCGCCCGCCGCC
TGGGCCCCGAGCCTTCTGCGCCGCCCGGGTGCGTCCCGCCACCCTCGGAGGACGGCCGGCCATGGACGCCTGCAAGT
TGGAGCCGAGCGGGAGGGTGTGAGCGGGCCGGGGCCAGGAGCCCGCGCCGCGCAACCGGGCAGCCGGGCGCGCCGGG
GGTGGGTCCCTCTCCCCAGCCCCGCTCTGCGTGGAAGAAGAGGGCGGGGACCGGCGCCGGGAGGAGAGCGGAGGAGG
CGAAGGGGCATGACTCGTGCAACTTGCGGCGGGCATCTGCCGAGCCTCTGAGCCGGCGGCGGCCCGGGGCCCGGACT
GCGGCCGCGCGGATCCACCCAGCCCACCCCGCCCCGGCCGACGGCTGCAGCTGACCTGGATCCTTCGAGCGCCCGCC
GACCGCCAGCGATCTTCCCTCATCTTCCGGGCTGGTTTCTGCTGCGCGAGGAGCGCTGCCCTCGCCGCCCCTCTCGC
CGGACCCCCGGCCCCCGATGGCTCGGATGGGGCTTGCGGGCGCCGCTGGACGCTGGTGGGGACTCGCTCTCGGCTTG
ACCGCATTCTTCCTCCCAGGCGTCCACTCCCAGGTGGTCCAGGTGAACGACTCCATGTATGGCTTCATCGGCACAGA
CGTGGTTCTGCACTGCAGCTTTGCCAACCCGCTTCCCAGCGTGAAGATCACCCAGGTCACATGGCAGAAGTCCACCA
ATGGCTCCAAGCAGAACGTGGCCATCTACAACCCATCCATGGGCGTGTCCGTGCTGGCTCCCTACCGCGAGCGTGTG
GAATTCCTGCGGCCCTCCTTCACCGATGGCACTATCCGCCTCTCCCGCCTGGAGCTGGAGGATGAGGGTGTCTACAT
CTGCGAGTTTGCTACCTTCCCTACGGGCAATCGAGAAAGCCAGCTCAATCTCACGGTGATGGCCAAACCCACCAATT
GGATAGAGGGTACCCAGGCAGTGCTTCGAGCCAAGAAGGGGCAGGATGACAAGGTCCTGGTGGCCACCTGCACCTCA

GCCAATGGGAAGCCTCCCAGTGTGGTATCCTGGGAAACTCGGTTAAAAGGTGAGGCAGAGTACCAGGAGATCCGGAA
CCCCAATGGCACAGTGACGGTCATCAGCCGCTACCGCCTGGTGCCCAGCAGGGAAGCCCACCAGCAGTCCTTGGCCT
GCATCGTCAACTACCACATGGACCGCTTCAAGGAAAGCCTCACTCTCAACGTGCAGTATGAGCCTGAGGTAACCATT
GAGGGGTTTGATGGCAACTGGTACCTGCAGCGGATGGACGTGAAGCTCACCTGCAAAGCTGATGCTAACCCCCCAGC
CACTGAGTACCACTGGACCACGCTAAATGGCTCTCTCCCCAAGGGTGTGGAGGCCCAGAACAGAACCCTCTTCTTCA
AGGGACCCATCAACTACAGCCTGGCAGGGACCTACATCTGTGAGGCCACCAACCCCATCGGTACACGCTCAGGCCAG
GTGGAGAATTCCCCTACACCCCGTCTCCTCCCGAACATGGGCGGCGCGCCGGGCCGGTGCCCACGGCCATCATTGGG
GGCGTGGCGGGGAGCATCCTGCTGGTGTTGATTGTGGTCGGCGGGATCGTGGTCGCCCTGCGTCGGCGCCGGCACAC
CTTCAAGGGTGACTACAGCACCAAGAAGCACGTGTATGGCAACGGCTACAGCAAGGCAGGCATCCCCCAGCACCACC
CACCAATGGCACAGAACCTGCAGTACCCCGACGACTCAGACGACGAGAAGAAGGCCGGCCCACTGGGTGGAAGCAGC
TATGAGGAGGAGGAGGAGGAGGAGGAGGGCGGTGGAGGGGGCGAGCGCAAGGTGGGCGGCCCCCACCCCAAATATGA
CGAGGACGCCAAGCGGCCCTACTTCACCGTGGATGAGGCCGAGGCCCGTCAGGACGGCTACGGGGACCGGACTCTGG
GCTACCAGTACGACCCTGAGCAGCTGGACTTGGCTGAGAACATGGTTTCTCAGAACGACGGGTCTTTCATTTCCAAG
AAGGAGTGGTACGTGTAGCCCCCCTTCCAGAGCCTCTGTCTGTGACCGCTCCTAACCAGCCCCTCCCCGCACGCCCC
CTGCCCACCCCCCACCTCCCACTCCAGGAGCTGAACAGAGACTTGCCCAGCTGCCCAAAGCCAGCCCCGAACTCCTG
GGGGGCCAGGGGAGCCCAGGGCAGCCACGACTTGGCTTTGTGTTTTATTTCCTC

<210> SEQ ID NO 51

<211> Length : 3,125

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 51

>HUMPHOSLIP_PEA_2_T6

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGGAGGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCG
GACCTGCGGGGCAAAGAAGGCCACTTCTACTACAACATCTCTGAGGTGAAGGTCACAGAGCTGCAACTGACATCTTC
CGAGCTCGATTTCCAGCCACAGCAGGAGCTGATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGAC
AGCTGCTCTACTGGTTCTTCTATGATGGGGGCTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTG
GAGCTCTCCCGGGATCCCGCTGGACGGATGAAAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGC
GGCCTTCGGGGGAACCTTCAAGAAGGTGTATGATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCA
ACCAGCAGATCTGCCCTGTCCTCTACCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGC

AGTTCTGTGGACGAGCTTGTTGGCATTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACAT
GGACTTCCGGGGGGCCTTCTTCCCCCTGACTGAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGC
AGGAGGAAGAGCGGATGGTGTATGTGGCCTTCTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCG
GGGGCCCTGCAGCTGTTGCTGGTGGGGGACAAGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGG
GAGCATTGTCCTGCTGAGCCCAGCAGTGATTGACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCT
GCACCATCAAGCCCTCTGGCACCACCATCTCTGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCT
GAGGTCCAGCTGTCCAGCATGACTATGGACGCCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCAC
GCAGCTGGACCTGCGCAGGTTCCGAATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGG
CCCCTCTGAAGACCATGCTGCAGATTGGGGTGATGCCCATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCA
CTACCTGAGGGCATCAACTTTGTGCATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCA
CTTTGCCAAAGGGCTGCGAGAGGTGATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGT
CCACAGCAGCTGTCTGAGCCCTCAATCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTT
TTCCCACATTCATAGCCTGTAGTGCCCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAA
TTTAATTCCATAATCAATCTATCAATTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTC
CTGGGATGGAATCAGTGCATCATAAAGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCT
GCGGGAATAAAGTGCTAACTTGCCCCCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTG
GGAAGGGGGCAGTCCTGGCGGCAGAACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAG
GCTGGGAGGGGCTGCATCAGGCCGTGGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGA
ACATGGTGAAGGCAGCGAGGGGCTTGTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCC
TGACCCCACCATTCCCTCCTCCCCATATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCC
CAGAACCCTCCCCACCCCTCTCCAAAAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAG
GCAGCCAGAGGGAAAGGACGGGTGGGTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTA
CCTTGATCGTGAGAAAGGCGATGTGGGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTC
ACTAACCAGGGCCGGCGCTGCACCTCCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTC
ACAGCTACCAGCAAGACCTTAGGGCTGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTG
GCCACTCCCAGGCATACCCGCTCCCAATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTC
TGGTTGAGGGAATCCACAAACCACTCATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGAT
CTGGTATTTCTAAAGCAGGATGGGGTAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAA
GGCCTACCCCTCAGGAAATTCAAGGGGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTA
CTGCATGTCCAGTATTCTACTAAATGTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 52

<211> Length : 3,263

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 52

>HUMPHOSLIP_PEA_2_T7

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT
ACAACATCTCTGAGCCTGGACTTGAAAGGGGAGCAGACAAATTTCCTGTCGTTGGGGGAAGTTCCCTCTTCTTGGCC
CTGGATCTGACCCTGAGGCCTCCTGTAGGGTGAAGGTCACAGAGCTGCAACTGACATCTTCCGAGCTCGATTTCCAG
CCACAGCAGGAGCTGATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGACAGCTGCTCTACTGGTT
CTTCTATGATGGGGGCTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTGGAGCTCTCCCGGGATC
CCGCTGGACGGATGAAAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGCGGCCTTCGGGGGAACC
TTCAAGAAGGTGTATGATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAGATCTGCCC
TGTCCTCTACCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGCAGTTCTGTGGACGAGC
TTGTTGGCATTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCCGGGGGGCC
TTCTTCCCCCTGACTGAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGAT
GGTGTATGTGGCCTTCTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGT
TGCTGGTGGGGGACAAGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTGTCCTGCTG
AGCCCAGCAGTGATTGACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTC
TGGCACCACCATCTCTGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCAGCTGTCCA
GCATGACTATGGACGCCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGC
AGGTTCCGAATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTGAAGACCAT
GCTGCAGATTGGGGTGATGCCCATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGCATCA
ACTTTGTGCATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCAAAGGGCTG
CGAGAGGTGATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCAGCTGTCTG
AGCCCTCAATCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACATTCATAGC
CTGTAGTGCCCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTCCATAATCA
ATCTATCAATTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGT
GCATCATAAAGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCT
AACTTGCCCCCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCT
GGCGGCAGAACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCA
TCAGGCCGTGGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGC
GAGGGGCTTGTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCACCATTCCC
TCCTCCCCATATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCCTCCCCACC
CCTCTCCAAAAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGGAAAG
GACGGGTGGGTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATCGTGAGAAA
GGCGATGTGGGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCAGGGCCGGC

GCTGCACCTCCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGA
CCTTAGGGCTGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCCCAGGCATA
CCCGCTCCCAATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAGGGAATCCA
CAAACCACTCATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATTTCTAAAGC
AGGATGGGGTAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACCCCTCAGGA
AATTCAAGGGGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGTCCAGTATT
CTACTAAATGTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 53

<211> Length : 3,256

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 53

>HUMPHOSLIP_PEA_2_T14

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT
ACAACATCTCTGAGGTGAAGGTCACAGAGCTGCAACTGACATCTTCCGAGCTCGATTTCCAGCCACAGCAGGAGCTG
ATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGACAGCTGCTCTACTGGTTCTTCTATGATGGGGG
CTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTGGAGCTCTCCCGGGATCCCGCTGGACGGATGA
AAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGCGGCCTTCGGGGGAACCTTCAAGAAGGTGTAT
GATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAGGTGTGGGCAGCGACAGGTCGCAG
GGTGGCAAGGGTGGGCATGCTCTCACTTTGAGAAGGCCCTGACTCTGGCTCCCACCTCGCAGATCTGCCCTGTCCTC
TACCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGCAGTTCTGTGGACGAGCTTGTTGG
CATTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCCGGGGGGCCTTCTTCC
CCCTGACTGAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGATGGTGTAT
GTGGCCTTCTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGTTGCTGGT
GGGGGACAAGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTGTCCTGCTGAGCCCAG
CAGTGATTGACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTCTGGCACC
ACCATCTCTGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCAGCTGTCCAGCATGAC
TATGGACGCCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGCAGGTTCC
GAATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTGAAGACCATGCTGCAG

ATTGGGGTGATGCCCATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGCATCAACTTTGT
GCATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCAAAGGGCTGCGAGAGG
TGATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCAGCTGTCTGAGCCCTC
AATCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACATTCATAGCCTGTAGT
GCCCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTCCATAATCAATCTATC
AATTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGTGCATCAT
AAAGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCTAACTTGC
CCCCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCTGGCGGCA
GAACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCATCAGGCC
GTGGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGCGAGGGGC
TTGTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCACCATTCCCTCCTCCC
CATATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCCTCCCCACCCCTCTCC
AAAAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGGAAAGGACGGGT
GGGTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATCGTGAGAAAGGCGATG
TGGGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCAGGGCCGGCGCTGCAC
CTCCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGACCTTAGG
GCTGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCCCAGGCATACCCGCTC
CCAATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAGGGAATCCACAAACCA
CTCATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATTTCTAAAGCAGGATGG
GGTAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACCCCTCAGGAAATTCAA
GGGGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGTCCAGTATTCTACTAA
ATGTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 54

<211> Length : 3,164

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 54

>HUMPHOSLIP_PEA_2_T16

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT

ACAACATCTCTGAGGTGAAGGTCACAGAGCTGCAACTGACATCTTCCGAGCTCGATTTCCAGCCACAGCAGGAGCTG
ATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGACAGCTGCTCTACTGGTTCTTCTATGATGGGGG
CTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTGGAGCTCTCCCGGGATCCCGCTGGACGGATGA
AAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGCGGCCTTCGGGGGAACCTTCAAGAAGGTGTAT
GATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAGATCTGCCCTGTCCTCTACCACGC
AGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTTCTGTGGACGAGCTTGTTGGCATTGACTATTCCCT
CATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCCGGGGGGCCTTCTTCCCCCTGACTGAGAGGA
ACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGATGGTGTATGTGGCCTTCTCTGAG
TTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGTTGCTGGTGGGGGACAAGGTGCC
CCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTGTCCTGCTGAGCCCAGCAGTGATTGACTCCC
CATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTCTGGCACCACCATCTCTGTCACT
GCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCAGCTGTCCAGCATGACTATGGACGCCCGTCT
CAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGCAGGTTCCGAATCTATTCCAACC
ATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTGAAGACCATGCTGCAGATTGGGGTGATGCCC
ATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGCATCAACTTTGTGCATGAGGTGGTGAC
GAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCAAAGGGCTGCGAGAGGTGATTGAGAAGAACC
GGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCAGCTGTCTGAGCCCTCAATCCCCAAGCTGGC
AGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACATTCATAGCCTGTAGTGCCCCCTCTAACCCC
CAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTCCATAATCAATCTATCAATTACAGTCCGTCC
ACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGTGCATCATAAAGGGCATTCTTTA
AGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCTAACTTGCCCCCAGGCTGTCTAT
GGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCTGGCGGCAGAACCCGGCCTGCAG
GGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCATCAGGCCGTGGAGCTGGTTGCT
GTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGCGAGGGGCTTGTCGGTGGGAACC
ATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCACCATTCCCTCCTCCCCATATAACTGCTCAC
TCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCCTCCCCACCCCTCTCCAAAAAAACTTGCCCA
TACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGGAAAGGACGGGTGGGTCCTGCTCCTTT
GCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATCGTGAGAAAGGCGATGTGGGAGAACTCCTTC
ACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCAGGGCCGGCGCTGCACCTCCATCTGCCCCAC
CAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGACCTTAGGGCTGGGAATTCCTCC
ACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCCCAGGCATACCCGCTCCCAATCCTCCACAGC
AGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAGGGAATCCACAAACCACTCATCCCCCATGAA
ATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATTTCTAAAGCAGGATGGGGTAAAAATGAGGGG
TGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACCCCTCAGGAAATTCAAGGGGCCAAGCTAGAT
AACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGTCCAGTATTCTACTAAATGTTTTATCTGTGA
AAGTAGA

<210> SEQ ID NO 55

<211> Length : 2,886

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 55

>HUMPHOSLIP_PEA_2_T17

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT
ACAACATCTCTGAGAAGGTGTATGATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAG
ATCTGCCCTGTCCTCTACCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGCAGTTCTGT
GGACGAGCTTGTTGGCATTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCC
GGGGGGCCTTCTTCCCCCTGACTGAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAA
GAGCGGATGGTGTATGTGGCCTTCTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCT
GCAGCTGTTGCTGGTGGGGGACAAGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTG
TCCTGCTGAGCCCAGCAGTGATTGACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATC
AAGCCCTCTGGCACCACCATCTCTGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCA
GCTGTCCAGCATGACTATGGACGCCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGG
ACCTGCGCAGGTTCCGAATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTG
AAGACCATGCTGCAGATTGGGGTGATGCCCATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGA
GGGCATCAACTTTGTGCATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCA
AAGGGCTGCGAGAGGTGATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCA
GCTGTCTGAGCCCTCAATCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACA
TTCATAGCCTGTAGTGCCCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTC
CATAATCAATCTATCAATTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATG
GAATCAGTGCATCATAAAGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAAT
AAAGTGCTAACTTGCCCCCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGG
GCAGTCCTGGCGGCAGAACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAG
GGGCTGCATCAGGCCGTGGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTG
AAGGCAGCGAGGGGCTTGTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCA
CCATTCCCTCCTCCCCATATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCC
TCCCCACCCCTCTCCAAAAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAG

AGGGAAAGGACGGGTGGGTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATC
GTGAGAAAGGCGATGTGGGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCA
GGGCCGGCGCTGCACCTCCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTAC
CAGCAAGACCTTAGGGCTGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCC
CAGGCATACCCGCTCCCAATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAG
GGAATCCACAAACCACTCATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATT
TCTAAAGCAGGATGGGGTAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACC
CCTCAGGAAATTCAAGGGGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGT
CCAGTATTCTACTAAATGTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 56

<211> Length : 3,100

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 56

>HUMPHOSLIP_PEA_2_T18

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT
ACAACATCTCTGAGGTGAAGGTCACAGAGCTGCAACTGACATCTTCCGAGCTCGATTTCCAGCCACAGCAGGAGCTG
ATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGACAGCTGCTCTACTGGTTCTTGAAGGTGTATGA
TTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAGGTGTGGGCAGCGACAGGTCGCAGGG
TGGCAAGGGTGGGCATGCTCTCACTTTGAGAAGGCCCTGACTCTGGCTCCCACCTCGCAGATCTGCCCTGTCCTCTA
CCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGCAGTTCTGTGGACGAGCTTGTTGGCA
TTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCCGGGGGGGCCTTCTTCCCC
CTGACTGAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGATGGTGTATGT
GGCCTTCTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGTTGCTGGTGG
GGGACAAGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTGTCCTGCTGAGCCCAGCA
GTGATTGACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTCTGGCACCAC
CATCTCTGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCAGCTGTCCAGCATGACTA
TGGACGCCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGCAGGTTCCGA
ATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTGAAGACCATGCTGCAGAT

TGGGGTGATGCCCATGCTCAATGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGCATCAACTTTGTGC
ATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCAAAGGGCTGCGAGAGGTG
ATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCAGCTGTCTGAGCCCTCAA
TCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACATTCATAGCCTGTAGTGC
CCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTCCATAATCAATCTATCAA
TTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGTGCATCATAA
AGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCTAACTTGCCC
CCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCTGGCGGCAGA
ACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCATCAGGCCGT
GGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGCGAGGGGCTT
GTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCACCATTCCCTCCTCCCCA
TATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCCTCCCCACCCCTCTCCAA
AAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGGAAAGGACGGGTGG
GTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATCGTGAGAAAGGCGATGTG
GGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCAGGGCCGGCGCTGCACCT
CCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGACCTTAGGGC
TGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCCCAGGCATACCCGCTCCC
AATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAGGGAATCCACAAACCACT
CATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATTTCTAAAGCAGGATGGGG
TAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACCCCTCAGGAAATTCAAGG
GGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGTCCAGTATTCTACTAAAT
GTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 57

<211> Length : 3,254

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 57

>HUMPHOSLIP_PEA_2_T19

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA
GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC
GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA
CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGAGCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTG
CTGGCAGGCGCACATGCAGAGTTCCCAGGCTGCAAGATCCGCGTCACCTCCAAGGCGCTGGAGCTGGTGAAGCAGGA
GGGGCTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCACCATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACT

ACAACATCTCTGAGGTGAAGGTCACAGAGCTGCAACTGACATCTTCCGAGCTCGATTTCCAGCCACAGCAGGAGCTG
ATGCTTCAAATCACCAATGCCTCCTTGGGGCTGCGCTTCCGGAGACAGCTGCTCTACTGGTTCTTCTATGATGGGGG
CTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTGGAGCTCTCCCGGGATCCCGCTGGACGGATGA
AAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGCGGCCTTCGGGGGAACCTTCAAGAAGGTGTAT
GATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTCCTCAACCAGCAGATCTGCCCTGTCCTCTACCACGC
AGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTGTGCGCAGTTCTGTGGACGAGCTTGTTGGCATTGACT
ATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGGACTTCCGGGGGGCCTTCTTCCCCCTGACT
GAGAGGAACTGGAGCCTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGATGGTGTATGTGGCCTT
CTCTGAGTTCTTCTTCGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGTTGCTGGTGGGGGACA
AGGTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGGAGCATTGTCCTGCTGAGCCCAGCAGTGATT
GACTCCCCATTGAAGCTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTCTGGCACCACCATCTC
TGTCACTGCTAGCGTCACCATTGCCCTGGTCCCACCAGACCAGCCTGAGGTCCAGCTGTCCAGCATGACTATGGACG
CCCGTCTCAGCGCCAAGATGGCTCTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGCAGGTTCCGAATCTAT
TCCAACCATTCTGCACTGGAGTCGCTGGCTCTGATCCCATTACAGGCCCCTCTGAAGACCATGCTGCAGATTGGGGT
GATGCCCATGCTCAATGGTAAGGCTGGGGTGTGAGGATGGAGGAAGAAAGGAGGGGTGAACTGGGCGGGCCCAGACT
GAGCGGGGTGCTCCCACCCACAGAGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGCATCAACTTTGTGC
ATGAGGTGGTGACGAACCATGCGGGATTCCTCACCATCGGGGCTGATCTCCACTTTGCCAAAGGGCTGCGAGAGGTG
ATTGAGAAGAACCGGCCTGCTGATGTCAGGGCGTCCACTGCCCCCACACCGTCCACAGCAGCTGTCTGAGCCCTCAA
TCCCCAAGCTGGCAGCTGTCATTCAGGACCCCAACCCCTCTCAGCCCCTCTTTTCCCACATTCATAGCCTGTAGTGC
CCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGATTTGAAGCTGTACCCAATTTAATTCCATAATCAATCTATCAA
TTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGTGCATCATAA
AGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCTAACTTGCCC
CCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCTGGCGGCAGA
ACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCATCAGGCCGT
GGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGCGAGGGGCTT
GTCGGTGGGAACCATGTGGCCGGCGCCCTGAGGAGCAATGTTCGTGAGCTCCTGACCCCACCATTCCCTCCTCCCCA
TATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAACCCTTTATTCTTCCCAGAACCCTCCCCACCCCTCTCCAA
AAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGACCCAGCCTAAAAGGCAGCCAGAGGGAAAGGACGGGTGG
GTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCAGGCCCAGTCCCTACCTTGATCGTGAGAAAGGCGATGTG
GGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCACTAACCAGGGCCGGCGCTGCACCT
CCATCTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGACCTTAGGGC
TGGGAATTCCTCCACACTTGCCCTCTGTGGGCCAGAGCCAGGCAGCCAGCTGGCCACTCCCAGGCATACCCGCTCCC
AATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGAATCTCTACCTTACCTTCTGGTTGAGGGAATCCACAAACCACT
CATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATAGGATCTGGTATTTCTAAAGCAGGATGGGG
TAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGTTCAAGGCCTACCCCTCAGGAAATTCAAGG
GGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGACTTACTGCATGTCCAGTATTCTACTAAAT
GTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 58

<211> Length : 1,533

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 58

>AI076020_T0

CGCTCAGTCCGGCAGCGCAGCAGGAGGGAGGCGCGAGGCAGGAGCCGGCGGCTGGGCTCCGCAGCGCAGCCAGCGCA
GCGCGGCGCCCCGGGCCCGGCCCCATGCCCGCAGCCCCGCCGGACCGTCCTTGAGCGCGGGCGCCTAGCCCGCGCCC
CCTGCCCGCCGGCACCATTGCCCCGACGGCGCGGCCGGGCGGCCCGGCGCTCCCCAGGCTCCGCGCGGGCCGAAAGA
CGCTGCTAGCGGCCGCCGCGGGTGTGGTGATGCTGCTGGTGCTGGTGGTGCTCATCCCCGTGCTGGTGAGCTCGGGC
GGCCCGGAAGGCCACTATGAGATGCTGGGCACCTGCCGCATGGTGTGCGACCCCTACCCCGCGCGGGGCCCCGGCGC
CGGCGCGCGGACCGACGGCGGCGACGCCCTGAGCGAGCAGAGCGGCGCGCCCCCGCCTTCCACGCTGGTGCAGGGCC
CCCAGGGGAAGCCGGGCCGCACCGGCAAGCCCGGCCCTCCGGGGCCTCCCGGGGACCCAGGTCCTCCCGGCCCTGTG
GGGCCGCCGGGGGGAGAAGGGTGAGCCAGGCAAGCCGGGCCCTCCGGGGCTGCCGGGCGCGGGGGGCAGCGGCGCCAT
CAGCACTGCCACCTACACCACGGTGCCGCGCGTGGCCTTCTACGCCGGCCTCAAGAACCCCCACGAGGGTTACGAGG
TACTCAAGTTTGACGACGTGGTCACCAACCTAGGCAACAACTACGACGCGGCCAGCGGCAAGTTTACGTGCAACATT
CCCGGCACCTACTTTTTCACCTACCATGTCCTCATGCGCGGCGGCGACGGCACCAGTATGTGGGCAGACCTCTGCAA
GAATGGCCAGGTGCGGGCCAGTGCTATTGCCCAGGACGCGGACCAGAACTACGACTACGCCAGCAACAGCGTGATCC
TGCACCTGGACGCCGGCGACGAGGTCTTCATCAAGCTGGATGGAGGCAAAGCACACGGCGGCAACAGCAACAAATAC
AGCACGTTCTCTGGCTTCATCATCTACTCCGACTGAGCTCCCCACGTCTCCCTCCACCCACGTCCCTCACCCGCCGG
GGTCCCCTCCGGGCGGGGCAGACGATGACTCGCCCCTCGCCCACCCGCTCGCTGCCCGGCCCTCCCCGGCTATGACG
CCCCCGGCCCGTGCTCAACACCGCCTGGGCCACAGCTAGGCCCTCCCACCGGCTCGCTGCAGAGCCGGGCCCAGCGC
GCCCTGTCCCCGTGCCAGGGAACCGGGGTTGACCGCCCCCGCCCAGCCCGCGCTATATATTTGTACAATAGGACTGT
TTACTGCCCACCTCCGCCTGCCAGCCCACCCCAGCCTGGGGAGAGGTCGCGGCGGCGGGTTTGCTTCCTGCGCTCTG
AGATGAGCTGCCCTCGGCTCCCTCCGGGGTGGCGCGCCCGGGGGAGGGGGGAGTTGGGGGCTGGATAGCTTCCCAGC
ACCCTCAGAGCCCCCGCCCGGCTGTGCCCCGTCTGACCAAAGTTATAATAAAAACATTTTCACCCCGCAG

<210> SEQ ID NO 59

<211> Length : 6,659

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 59

>M79217_PEA_1_T1

TGATGTCTGCTAATGGAAGATAAATGAGAAGCAAACTGGGAAATACATTTTGTCTCAGGATTACTGCATCTACTACT
GGATAAAGATCAAAAGAGTTCTCTTCAACCCTTTCAACTCTACATTTAACAAATTGAGCTTTTCAGAGTCTTTTTTT
GTAAAGTATTTCCAAAGAAGACTTATAGGTTAGGAATAAACATAAACTACCCAGGTTGGCTAGGAAGGTATTTCTGT
TCATCTAAAGATGATGCCCAGGTGTGGAACAGGATAAGAAAAGACCATGGACATCTTTGTCCCATGAATTTAGTTGG
TCATCGTGTTACAGGGCTATAATGCCGCTCTAGATCCAGTTAAATAAGAAGTGGGGAGGGGTTGTAAACTGCAGCTT
TTTGGGGCACTTATCCATTTATTACCCCAAGTAAAAGACCTATACCAAACAGCAAACAACATCTCTGCATTGTCATT
ATAATGTTCTTTGAGACACAGCCAGTGTTCCAGCCATTGTTCCATCTAAGATTTAAGCATTTTCTAGAAATGTATGG
TGGCAGGGGTGTTGAACATAACTTCTTCAAGACTGACATGGTTCTCTTTCTTTTGCAGGCCTGATTGTTGGCAAAGG
CATCATAAGAAGCTGGCATTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATAGACACTATGCATATTTGTT
GGTCAGCAAAACCAAGAAACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAGGGTGTTTTTCCTGGGTTTC
ATCATCAGGTACCTCCTCCCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTGATAAAGATTAAGGACATGT
TCTTTGGTCAACAGCCAGAACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAGCTGCAGCTGAGGAAAATGA
AATGTTCATTTTATTTGGTGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTGTCAGTGAAACAGAGATCGT
TTTGTGGAATAGCTAACCCATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGCAGGCTGCAGAGGACTCATGAC
AGGCTATACCATGCTGCGGAATGGGGGCGCGGGGAACGGAGGTCAGACCTGCATGCTGCGCTGGTCCAACCGCATCC
GCCTCACGTGGCTCAGCTTCACGCTCTTTGTCATCCTGGTCTTCTTCCCGCTCATCGCCCACTATTACCTCACCACT
CTGGATGAGGCTGATGAGGCAGGCAAGCGGATTTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACGT
GCTGGATCTGTGCCGCATCCGGGAGTCGGTGAGTGAAGAGCTCCTGCAGCTGGAGGCCAAGCGCCAAGAGCTGAACA
GCGAGATCGCCAAGCTGAATCTGAAGATCGAAGCCTGTAAGAAGAGCATTGAGAACGCCAAGCAGGACCTGCTCCAG
CTCAAGAATGTCATCAGCCAGACCGAGCATTCCTACAAGGAGCTCATGGCCCAGAACCAGCCCAAGCTGTCCCTGCC
CATCCGACTGCTCCCAGAGAAGGACGATGCCGGCCTCCCTCCCCCGAAGGCCACTCGGGGCTGCCGGCTACACAACT
GCTTTGATTATTCTCGTTGCCCTCTCACCTCTGGCTTCCCGGTCTACGTCTATGACAGTGACCAGTTTGTCTTTGGC
AGCTACCTGGATCCCTTGGTCAAGCAGGCTTTTCAGGCGACAGCACGAGCTAACGTTTATGTTACAGAAAATGCAGA
CATCGCCTGCCTTTACGTGATACTAGTGGGAGAGATGCAGGAGCCGGTGGTGCTGCGGCCTGCTGAGCTGGAGAAGC
AGTTGTATTCCCTGCCACACTGGCGGACGGATGGACACAACCATGTCATCATCAATCTGTCACGTAAGTCAGATACA
CAGAACCTTCTCTATAACGTCAGTACTGGCCGTGCCATGGTGGCCCAGTCCACCTTCTACACTGTCCAGTACAGACC
TGGCTTTGACTTGGTCGTATCACCGCTGGTCCATGCCATGTCTGAGCCCAACTTCATGGAAATCCCACCACAGGTGC
CGGTGAAGCGGAAATATCTCTTCACCTTCCAGGGCGAGAAGATTGAGTCTCTGAGGTCTAGCCTTCAGGAGGCCCGC
TCCTTCGAAGAGGAAATGGAGGGCGACCCTCCCGCCGACTACGATGACCGGATCATTGCCACCCTGAAGGCGGTGCA
GGACAGCAAGCTGGATCAGGTCCTGGTGGAATTCACCTGCAAAAACCAGCCCAAACCCAGCCTGCCAACTGAGTGGG
CACTGTGTGGAGAGCGGGAGGACCGCTTGGAATTGCTGAAGCTCTCCACCTTCGCCCTCATCATTACCCCCGGGGAC
CCTCGCTTGGTTATTTCCTCTGGGTGTGCAACACGGCTCTTCGAAGCCCTGGAAGTCGGTGCCGTCCCGGTGGTGCT
GGGGGAGCAGGTCCAGCTTCCCTACCAGGACATGCTGCAGTGGAACGAGGCGGCCCTGGTGGTGCCAAAGCCTCGTG
TTACCGAGGTTCATTTCCTGCTCAGAAGCCTCTCCGATAGTGACCTCCTGGCTATGAGGCGGCAAGGCCGCTTTCTC

TGGGAGACTTACTTCTCCACTGCTGACAGTATTTTTAATACCGTGCTGGCTATGATTAGGACTCGCATCCAGATCCC
AGCCGCTCCCATCCGGGAAGAGGCGGCAGCTGAGATCCCCCACCGTTCAGGCAAGGCGGCTGGAACTGACCCCAACA
TGGCTGACAACGGGGACCTGGACCTGGGGCCAGTGGAGACGGAGCCGCCCTACGCCTCACCCAGATACCTCCGCAAT
TTCACTCTGACTGTCACTGACTTTTACCGCAGCTGGAACTGTGCTCCAGGGCCTTTCCATCTTTTCCCCCACACTCC
CTTTGACCCTGTGTTGCCCTCAGAGGCCAAATTCTTGGGCTCAGGGACTGGCTTTCGGCCTATTGGTGGTGGAGCTG
GGGGTTCTGGCAAGGAATTTCAGGCAGCGCTTGGAGGCAATGTTCCCCGAGAGCAGTTCACGGTGGTGATGTTGACT
TATGAGCGGGAGGAAGTGCTTATGAACTCTTTAGAGAGGCTGAATGGCCTCCCTTACCTGAACAAGGTCGTGGTGGT
GTGGAATTCTCCCAAGCTGCCATCAGAGGACCTTCTGTGGCCTGACATTGGCGTCCCCATCATGGTGGTCCGTACTG
AGAAGAACAGTTTGAACAACCGATTCTTACCCTGGAATGAAATTGAGACAGAGGCCATCCTGTCCATTGATGACGAT
GCTCACCTCCGCCATGACGAAATCATGTTTGGGTTCCGGGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCC
TGGCCGTTACCACGCATGGGACATCCCCCATCAGTCCTGGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGG
TGCTGACAGGTGCTGCCTTCTTTCACAAGTATTATGCCTACCTGTATTCTTATGTGATGCCCCAGGCCATCCGGGAC
ATGGTGGATGAATACATCAACTGTGAGGACATTGCCATGAACTTCCTTGTCTCCCACATCACTCGGAAGCCCCCCAT
CAAGGTGACCTCACGGTGGACATTCCGATGCCCAGGATGCCCTCAGGCCCTGTCTCATGATGACTCCCACTTCCACG
AGCGGCACAAGTGCATCAACTTCTTCGTGAAGGTGTACGGCTACATGCCCCTCCTGTACACGCAGTTCAGGGTGGAT
TCTGTGCTCTTCAAGACACGCCTGCCCCATGACAAGACCAAGTGCTTCAAGTTCATCTAGGGGCAGCGCACGGTCTG
GGGAAGAGGATGAGCAGAGGGAGGAAGATGGCTCCCAAGGTTCCTAGGCATTGCAGGACCTTGGGCACATCTGCTGG
TGGGTGGCCCAGAGCCTCTGCTGGAAGGGGCAGCAGGAGGAGTGGAAGGAAACCGCTGCCTTTATCTTGAAGTCAGC
CACACTGGGCCTGGAGCCCTGGGCGGAGTCCCCGGGGTTCCCCACACAGGGCACTGACTGATAGCTTACACTGAGGA
CTGTGGCGACTCTGCAGAGTCACTCACACCGTTCGTACGCCCAGGACAGCTGGTTCGTGGTTTTTACATTCAATAAC
AACTATTATGATTATTTAAAAAGAGAAAGTTTCAGATTTGCCATTCAAGGCTTATTTATATATATGTGTGTGTATAT
AAATACATGCACACACTTGCATACATATATATTTTTGGCTGGGGGAGTGTGAGTTTTGCCTTTCTAAGGGAGGGACC
GCGCAGGCTCCTTTGTTCTGTATTCTGGCGGAGATGGGTCCTGGCCTTGTGTCACTGGCTTATCCTTAAAGATCATC
TCCCATCCTCCCCAGCGCCATCTGTGTGCAGCAACCAGAAAGGGATGAACTTGGCCCTCTTGCGGGCCTGGACAAGG
TCTCTTCCTTACCCTTTCTGTTGCCAGTCAGCAACCTGTAACTCACATTCTCTTCCCAGTGAATCCCTGGGAGCGCC
TGACCCTGGTGGGCTGTTCAGCTTCCTGCTGCTGGGGCCAGCGATTTTTGAGGATTTATCTTTAGGCCAGGCTTGCC
TCCGTACTTATCCCTGCTCTCCCATTTCTCTCTTGTTTGAGAGAGAATGAGGAAGCAAAGAGTGAGAAAGAATAGGG
GCTGAAGACGCCACTCCCAGATGGCTCTTTCTATCCTGCTCTTCTGTTGAAACACACGTGCTGTGGGCCTCAGGCGT
TTCTGAAGTGCTCTTTCTTGGATTGGACAGGAGATCAGCAGCGTGCACATCTGCTGTGGTCTGAAGTGGTTTGCAGG
TCAGCCTCCTCTCCCTAGTGTAGAGCAAGCCAGTGTCCTTCGAGGAACCCACCCGGCTGGCCGGGAAGTTTTACAGC
AAGGCGCCTGCCTTGGGATAATTCCTTGGTGAAATTCACCTTCCCCCCGCCTCTGTCTGGAGCCCCATCCTGTGTTA
TCTGTGGTTTTTGGACCCCTAATGTCAGCTTGGCTGTAGGACTCCCCGAGGTTTGGTATGTGCTAGAACAATGGGAG
GCTGTGATTTGCTGTGTAAGCTCACATCCAGCCTTGGAATCTAACGGGCATTCACAACCCGAGTTACCACTTTCCAC
TCCCTGCTTAGGATTCTGTTCCCTGGGCTGAAACTGAAATAAGCTAATTTTTTGGGTCACGGTGGCAGTAGGGGAAC
CTAGGAGGGTGTGAGTGGCATTTGTCAGGGATTTAGCCCATGACGTGTTTCTTGAACCCTACTTTCTGGAAGTGGAG
TTGACTCTGGAAGTTTTCTAGCAACTGAACAAAAGCTCAGGTTTGTCCTGGTCATGCACATGCCTTAAGCCAGTTCC
GTCTTCCCTAGACCTTGGCATCCTGTGCTTCTATTTCTTGGAATACGTTCTCCTCTGACCTGCCTGTACCACGTGGG
TCCTCTTCAAGTACTGTTTTGAAGCTGGGCTCTTTTGTGTAGCTCCCACCCACCTGTAGGGCTAGCTCGGCTTAAGG

GAACTCTCCCCATTGGCAAACCGGACCCGGCCGCCGCCAGGACTGTGTTTCCAAAGGTTCCCCGCCCCCAACCCCAG
CATCAGCCTGTAGCTCCCCTGCTGAGGCAGTGTGGTTATGTTCCCAGCAGTGGGGGTCAGACGCCCTTCCTCAGAAC
TTTCTAGTTGCCCTCTACCTGACTCCTGACTTGTATTCCTTTTAGCAGTAGCCTTCTTCCCTCGGGGAGCCAAAGAG
TGTGGTGTGTGGCGCTATATTGTGGCTGCTATTTCATCTGGTTTCTTTTAATGTGAGGAACTCACATACTGACTTCA
GTGGGACTCGGTGAGCCGGGGCCGTCTGTGTGGTGGGACCCCCTTTAGCGGGACTCAGTGAGCTGGGGCCGTCTGTG
TGGTGGAGCCAGGGCCTCTCCCTTTAGTGGAGCCAGGTTGTCGGGCCCCGAATGTCACTGGTGGATCTAAGAAGGGC
TGAGTGGTCTGACACCAAAACATGCCGCAGGGAGGGCTGTGGTGCCGGTGCTTCCAACAAGGACAGCCCTCCTTGAC
CCTGAAAGGAACACTGGCTTGAAGGACTGCAGACAGGCTCTGAGGGGCACGCCCTCCTCAGCGAGAGGCAGCAAGGT
GGCCACAGTGTCACTGGTCAGGTGCTTCTCACCACGGGAAAGCCGCCGACCTGTGACTCGCTTGAGATGGGAAAGCG
GCGCCACAGACCCCGGGTCTCCTTGGCTGTCTGTGGGCCGCCCCTGGCCACCTTGTCCTGGCTCGCAGGGTGCAGGA
GCGCCTCGTTCTCTGGGTGGCCGGCTTGCTGCTCCGGTTTGGGCTGTCTTACCATAACACCGTCCCAGGGCTCTGCA
GGCCACTGTGAGCGCTGGCTCCCTGGGCAGTGCTCCTCCGTGTGGACTGTGCCTCAGGCCAGGGCTCACCAGCTGGG
GTCCTGTCCGGAAGGATGGGATCTTTCTGGGAGCTGCGCCGGACAGAGTGGGGAGCTCCTAGTTTGTGGGGGGAAGC
TTTGATATCCATGCCACGTCCATCCACCCCACCCCTTTTCGTCACGAGCACAATGGTCTTACATTGGATTTTTGTAA
AAAAATAAAAATAAATGGAGACTTTAACTCAAGCAGC

<210> SEQ ID NO 60

<211> Length : 6,333

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 60

>M79217_PEA_1_T3

GGCTGCAGTGGCCGCCGCTGGAGACCGCGGGACCTGGCGCTGCAGCGAGATTTTCAAACACCGCATTCCCCTTAAAC
CTCCCAGGCTCCCATCACTCCCAGCAAAGAGCCAGCCCGCTACTTTATGGAGACAGGAGAGACTGTCAGACAGTTCC
TGCTGCTGCCCAGGTGCATCTGAGAGAGCAAGCCCTGGAGGTTCACTCTTTCAAGAAGTCGTGTGCTGAGGTGTAAT
GCTACACAAGTCAGAGGAAGGAAGGGTCCTGAAACACATGGCCTGATTGTTGGCAAAGGCATCATAAGAAGCTGGCA
TTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATAGACACTATGCATATTTGTTGGTCAGCAAAACCAAGAA
ACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAGGGTGTTTTTCCTGGGTTTCATCATCAGGTACCTCCTC
CCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTGATAAAGATTAAGGACATGTTCTTTGGTCAACAGCCAG
AACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAGCTGCAGCTGAGGAAAATGAAATGTTCATTTTATTTGG
TGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTGTCAGTGAAACAGAGATCGTTTTGTGGAATAGCTAACC
CATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGCAGGCTGCAGAGGACTCATGACAGGCTATACCATGCTGCG
GAATGGGGGCGCGGGGAACGGAGGTCAGACCTGCATGCTGCGCTGGTCCAACCGCATCCGCCTCACGTGGCTCAGCT
TCACGCTCTTTGTCATCCTGGTCTTCTTCCCGCTCATCGCCCACTATTACCTCACCACTCTGGATGAGGCTGATGAG
GCAGGCAAGCGGATTTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACGTGCTGGATCTGTGCCGCAT

CCGGGAGTCGGTGAGTGAAGAGCTCCTGCAGCTGGAGGCCAAGCGCCAAGAGCTGAACAGCGAGATCGCCAAGCTGA
ATCTGAAGATCGAAGCCTGTAAGAAGAGCATTGAGAACGCCAAGCAGGACCTGCTCCAGCTCAAGAATGTCATCAGC
CAGACCGAGCATTCCTACAAGGAGCTCATGGCCCAGAACCAGCCCAAGCTGTCCCTGCCCATCCGACTGCTCCCAGA
GAAGGACGATGCCGGCCTCCCTCCCCCGAAGGCCACTCGGGGCTGCCGGCTACACAACTGCTTTGATTATTCTCGTT
GCCCTCTCACCTCTGGCTTCCCGGTCTACGTCTATGACAGTGACCAGTTTGTCTTTGGCAGCTACCTGGATCCCTTG
GTCAAGCAGGCTTTTCAGGCGACAGCACGAGCTAACGTTTATGTTACAGAAAATGCAGACATCGCCTGCCTTTACGT
GATACTAGTGGGAGAGATGCAGGAGCCGGTGGTGCTGCGGCCTGCTGAGCTGGAGAAGCAGTTGTATTCCCTGCCAC
ACTGGCGGACGGATGGACACAACCATGTCATCATCAATCTGTCACGTAAGTCAGATACACAGAACCTTCTCTATAAC
GTCAGTACTGGCCGTGCCATGGTGGCCCAGTCCACCTTCTACACTGTCCAGTACAGACCTGGCTTTGACTTGGTCGT
ATCACCGCTGGTCCATGCCATGTCTGAGCCCAACTTCATGGAAATCCCACCACAGGTGCCGGTGAAGCGGAAATATC
TCTTCACCTTCCAGGGCGAGAAGATTGAGTCTCTGAGGTCTAGCCTTCAGGAGGCCCGCTCCTTCGAAGAGGAAATG
GAGGGCGACCCTCCCGCCGACTACGATGACCGGATCATTGCCACCCTGAAGGCGGTGCAGGACAGCAAGCTGGATCA
GGTCCTGGTGGAATTCACCTGCAAAAACCAGCCCAAACCCAGCCTGCCAACTGAGTGGGCACTGTGTGGAGAGCGGG
AGGACCGCTTGGAATTGCTGAAGCTCTCCACCTTCGCCCTCATCATTACCCCCGGGGACCCTCGCTTGGTTATTTCC
TCTGGGTGTGCAACACGGCTCTTCGAAGCCCTGGAAGTCGGTGCCGTCCCGGTGGTGCTGGGGGAGCAGGTCCAGCT
TCCCTACCAGGACATGCTGCAGTGGAACGAGGCGGCCCTGGTGGTGCCAAAGCCTCGTGTTACCGAGGTTCATTTCC
TGCTCAGAAGCCTCTCCGATAGTGACCTCCTGGCTATGAGGCGGCAAGGCCGCTTTCTCTGGGAGACTTACTTCTCC
ACTGCTGACAGTATTTTTAATACCGTGCTGGCTATGATTAGGACTCGCATCCAGATCCCAGCCGCTCCCATCCGGGA
AGAGGCGGCAGCTGAGATCCCCCACCGTTCAGGCAAGGCGGCTGGAACTGACCCCAACATGGCTGACAACGGGGACC
TGGACCTGGGGCCAGTGGAGACGGAGCCGCCCTACGCCTCACCCAGATACCTCCGCAATTTCACTCTGACTGTCACT
GACTTTTACCGCAGCTGGAACTGTGCTCCAGGGCCTTTCCATCTTTTCCCCCACACTCCCTTTGACCCTGTGTTGCC
CTCAGAGGCCAAATTCTTGGGCTCAGGGACTGGCTTTCGGCCTATTGGTGGTGGAGCTGGGGGTTCTGGCAAGGAAT
TTCAGGCAGCGCTTGGAGGCAATGTTCCCCGAGAGCAGTTCACGGTGGTGATGTTGACTTATGAGCGGGAGGAAGTG
CTTATGAACTCTTTAGAGAGGCTGAATGGCCTCCCTTACCTGAACAAGGTCGTGGTGGTGTGGAATTCTCCCAAGCT
GCCATCAGAGGACCTTCTGTGGCCTGACATTGGCGTCCCCATCATGGTGGTCCGTACTGAGAAGAACAGTTTGAACA
ACCGATTCTTACCCTGGAATGAAATTGAGACAGAGGCCATCCTGTCCATTGATGACGATGCTCACCTCCGCCATGAC
GAAATCATGTTTGGGTTCCGGGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCCTGGCCGTTACCACGCATG
GGACATCCCCCATCAGTCCTGGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGGTGCTGACAGGTGCTGCCT
TCTTTCACAAGTATTATGCCTACCTGTATTCTTATGTGATGCCCCAGGCCATCCGGGACATGGTGGATGAATACATC
AACTGTGAGGACATTGCCATGAACTTCCTTGTCTCCCACATCACTCGGAAGCCCCCCATCAAGGTGACCTCACGGTG
GACATTCCGATGCCCAGGATGCCCTCAGGCCCTGTCTCATGATGACTCCCACTTCCACGAGCGGCACAAGTGCATCA
ACTTCTTCGTGAAGGTGTACGGCTACATGCCCCTCCTGTACACGCAGTTCAGGGTGGATTCTGTGCTCTTCAAGACA
CGCCTGCCCCATGACAAGACCAAGTGCTTCAAGTTCATCTAGGGGCAGCGCACGGTCTGGGGAAGAGGATGAGCAGA
GGGAGGAAGATGGCTCCCAAGGTTCCTAGGCATTGCAGGACCTTGGGCACATCTGCTGGTGGGTGGCCCAGAGCCTC
TGCTGGAAGGGGCAGCAGGAGGAGTGGAAGGAAACCGCTGCCTTTATCTTGAAGTCAGCCACACTGGGCCTGGAGCC
CTGGGCGGAGTCCCCGGGGTTCCCCACACAGGGCACTGACTGATAGCTTACACTGAGGACTGTGGCGACTCTGCAGA
GTCACTCACACCGTTCGTACGCCCAGGACAGCTGGTTCGTGGTTTTTACATTCAATAACAACTATTATGATTATTTA
AAAAGAGAAAGTTTCAGATTTGCCATTCAAGGCTTATTTATATATATGTGTGTGTATATAAATACATGCACACACTT

GCATACATATATATTTTTGGCTGGGGGAGTGTGAGTTTTGCCTTTCTAAGGGAGGGACCGCGCAGGCTCCTTTGTTC
TGTATTCTGGCGGAGATGGGTCCTGGCCTTGTGTCACTGGCTTATCCTTAAAGATCATCTCCCATCCTCCCCAGCGC
CATCTGTGTGCAGCAACCAGAAAGGGATGAACTTGGCCCTCTTGCGGGCCTGGACAAGGTCTCTTCCTTACCCTTTC
TGTTGCCAGTCAGCAACCTGTAACTCACATTCTCTTCCCAGTGAATCCCTGGGAGCGCCTGACCCTGGTGGGCTGTT
CAGCTTCCTGCTGCTGGGGCCAGCGATTTTTGAGGATTTATCTTTAGGCCAGGCTTGCCTCCGTACTTATCCCTGCT
CTCCCATTTCTCTCTTGTTTGAGAGAGAATGAGGAAGCAAAGAGTGAGAAAGAATAGGGGCTGAAGACGCCACTCCC
AGATGGCTCTTTCTATCCTGCTCTTCTGTTGAAACACACGTGCTGTGGGCCTCAGGCGTTTCTGAAGTGCTCTTTCT
TGGATTGGACAGGAGATCAGCAGCGTGCACATCTGCTGTGGTCTGAAGTGGTTTGCAGGTCAGCCTCCTCTCCCTAG
TGTAGAGCAAGCCAGTGTCCTTCGAGGAACCCACCCGGCTGGCCGGGAAGTTTTACAGCAAGGCGCCTGCCTTGGGA
TAATTCCTTGGTGAAATTCACCTTCCCCCCGCCTCTGTCTGGAGCCCCATCCTGTGTTATCTGTGGTTTTTGGACCC
CTAATGTCAGCTTGGCTGTAGGACTCCCCGAGGTTTGGTATGTGCTAGAACAATGGGAGGCTGTGATTTGCTGTGTA
AGCTCACATCCAGCCTTGGAATCTAACGGGCATTCACAACCCGAGTTACCACTTTCCACTCCCTGCTTAGGATTCTG
TTCCCTGGGCTGAAACTGAAATAAGCTAATTTTTTGGGTCACGGTGGCAGTAGGGGAACCTAGGAGGGTGTGAGTGG
CATTTGTCAGGGATTTAGCCCATGACGTGTTTCTTGAACCCTACTTTCTGGAAGTGGAGTTGACTCTGGAAGTTTTC
TAGCAACTGAACAAAAGCTCAGGTTTGTCCTGGTCATGCACATGCCTTAAGCCAGTTCCGTCTTCCCTAGACCTTGG
CATCCTGTGCTTCTATTTCTTGGAATACGTTCTCCTCTGACCTGCCTGTACCACGTGGGTCCTCTTCAAGTACTGTT
TTGAAGCTGGGCTCTTTTGTGTAGCTCCCACCCACCTGTAGGGCTAGCTCGGCTTAAGGGAACTCTCCCCATTGGCA
AACCGGACCCGGCCGCCGCCAGGACTGTGTTTCCAAAGGTTCCCCGCCCCCAACCCCAGCATCAGCCTGTAGCTCCC
CTGCTGAGGCAGTGTGGTTATGTTCCCAGCAGTGGGGGTCAGACGCCCTTCCTCAGAACTTTCTAGTTGCCCTCTAC
CTGACTCCTGACTTGTATTCCTTTTAGCAGTAGCCTTCTTCCCTCGGGGAGCCAAAGAGTGTGGTGTGTGGCGCTAT
ATTGTGGCTGCTATTTCATCTGGTTTCTTTTAATGTGAGGAACTCACATACTGACTTCAGTGGGACTCGGTGAGCCG
GGGCCGTCTGTGTGGTGGGACCCCCTTTAGCGGGACTCAGTGAGCTGGGGCCGTCTGTGTGGTGGAGCCAGGGCCTC
TCCCTTTAGTGGAGCCAGGTTGTCGGGCCCCGAATGTCACTGGTGGATCTAAGAAGGGCTGAGTGGTCTGACACCAA
AACATGCCGCAGGGAGGGCTGTGGTGCCGGTGCTTCCAACAAGGACAGCCCTCCTTGACCCTGAAAGGAACACTGGC
TTGAAGGACTGCAGACAGGCTCTGAGGGGCACGCCCTCCTCAGCGAGAGGCAGCAAGGTGGCCACAGTGTCACTGGT
CAGGTGCTTCTCACCACGGGAAAGCCGCCGACCTGTGACTCGCTTGAGATGGGAAAGCGGCGCCACAGACCCCGGGT
CTCCTTGGCTGTCTGTGGGCCGCCCCTGGCCACCTTGTCCTGGCTCGCAGGGTGCAGGAGCGCCTCGTTCTCTGGGT
GGCCGGCTTGCTGCTCCGGTTTGGGCTGTCTTACCATAACACCGTCCCAGGGCTCTGCAGGCCACTGTGAGCGCTGG
CTCCCTGGGCAGTGCTCCTCCGTGTGGACTGTGCCTCAGGCCAGGGCTCACCAGCTGGGGTCCTGTCCGGAAGGATG
GGATCTTTCTGGGAGCTGCGCCGGACAGAGTGGGGAGCTCCTAGTTTGTGGGGGGAAGCTTTGATATCCATGCCACG
TCCATCCACCCCACCCCTTTTCGTCACGAGCACAATGGTCTTACATTGGATTTTTGTAAAAAAATAAAAATAAATGG
AGACTTTAACTCAAGCAGC

<210> SEQ ID NO 61

<211> Length : 6,297

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 61

>M79217_PEA_1_T8

GGCGCCCTTCGGCAAGTTCCGCAGTCGCCTGTCGGAAATGGCTGCCGGCCGGCAGGGGGAGCGGCGGATCAGGCGCG
GCCTGGAAGGCGGGCGGCCGGCAGCCAGAACGGCTTCTGGGACGCCGACTTTCGCGCAGGCGGCGGCGGCGGCGG
GCGGGTCCCTGAGCTGGAAGCCGGAGAGCAAGCCCTGGAGGTTCACTCTTTCAAGAAGTCGTGTGCTGAGGTGTAAT
GCTACACAAGTCAGAGGAAGGAAGGGTCCTGAAACACATGGCCTGATTGTTGGCAAAGGCATCATAAGAAGCTGGCA
TTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATAGACACTATGCATATTTGTTGGTCAGCAAAACCAAGAA
ACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAGGGTGTTTTTCCTGGGTTTCATCATCAGGTACCTCCTC
CCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTGATAAAGATTAAGGACATGTTCTTTGGTCAACAGCCAG
AACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAGCTGCAGCTGAGGAAAATGAAATGTTCATTTTATTTGG
TGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTGTCAGTGAAACAGAGATCGTTTTGTGGAATAGCTAACC
CATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGCAGGCTGCAGAGGACTCATGACAGGCTATACCATGCTGCG
GAATGGGGGCGCGGGGAACGGAGGTCAGACCTGCATGCTGCGCTGGTCCAACCGCATCCGCCTCACGTGGCTCAGCT
TCACGCTCTTTGTCATCCTGGTCTTCTTCCCGCTCATCGCCCACTATTACCTCACCACTCTGGATGAGGCTGATGAG
GCAGGCAAGCGGATTTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACGTGCTGGATCTGTGCCGCAT
CCGGGAGTCGGTGAGTGAAGAGCTCCTGCAGCTGGAGGCCAAGCGCCAAGAGCTGAACAGCGAGATCGCCAAGCTGA
ATCTGAAGATCGAAGCCTGTAAGAAGAGCATTGAGAACGCCAAGCAGGACCTGCTCCAGCTCAAGAATGTCATCAGC
CAGACCGAGCATTCCTACAAGGAGCTCATGGCCCAGAACCAGCCCAAGCTGTCCCTGCCCATCCGACTGCTCCCAGA
GAAGGACGATGCCGGCCTCCCTCCCCCGAAGGCCACTCGGGGCTGCCGGCTACACAACTGCTTTGATTATTCTCGTT
GCCCTCTCACCTCTGGCTTCCCGGTCTACGTCTATGACAGTGACCAGTTTGTCTTTGGCAGCTACCTGGATCCCTTG
GTCAAGCAGGCTTTTCAGGCGACAGCACGAGCTAACGTTTATGTTACAGAAAATGCAGACATCGCCTGCCTTTACGT
GATACTAGTGGGAGAGATGCAGGAGCCGGTGGTGCTGCGGCCTGCTGAGCTGGAGAAGCAGTTGTATTCCCTGCCAC
ACTGGCGGACGGATGGACACAACCATGTCATCATCAATCTGTCACGTAAGTCAGATACACAGAACCTTCTCTATAAC
GTCAGTACTGGCCGTGCCATGGTGGCCCAGTCCACCTTCTACACTGTCCAGTACAGACCTGGCTTTGACTTGGTCGT
ATCACCGCTGGTCCATGCCATGTCTGAGCCCAACTTCATGGAAATCCCACCACAGGTGCCGGTGAAGCGGAAATATC
TCTTCACCTTCCAGGGCGAGAAGATTGAGTCTCTGAGGTCTAGCCTTCAGGAGGCCCGCTCCTTCGAAGAGGAAATG
GAGGGCGACCCTCCCGCCGACTACGATGACCGGATCATTGCCACCCTGAAGGCGGTGCAGGACAGCAAGCTGGATCA
GGTCCTGGTGGAATTCACCTGCAAAAACCAGCCCAAACCCAGCCTGCCAACTGAGTGGGCACTGTGTGGAGAGCGGG
AGGACCGCTTGGAATTGCTGAAGCTCTCCACCTTCGCCCTCATCATTACCCCCGGGGACCCTCGCTTGGTTATTTCC
TCTGGGTGTGCAACACGGCTCTTCGAAGCCCTGGAAGTCGGTGCCGTCCCGGTGGTGCTGGGGGAGCAGGTCCAGCT
TCCCTACCAGGACATGCTGCAGTGGAACGAGGCGGCCCTGGTGGTGCCAAAGCCTCGTGTTACCGAGGTTCATTTCC
TGCTCAGAAGCCTCTCCGATAGTGACCTCCTGGCTATGAGGCGGCAAGGCCGCTTTCTCTGGGAGACTTACTTCTCC
ACTGCTGACAGTATTTTTAATACCGTGCTGGCTATGATTAGGACTCGCATCCAGATCCCAGCCGCTCCCATCCGGGA
AGAGGCGGCAGCTGAGATCCCCCACCGTTCAGGCAAGGCGGCTGGAACTGACCCCAACATGGCTGACAACGGGGACC
TGGACCTGGGGCCAGTGGAGACGGAGCCGCCCTACGCCTCACCCAGATACCTCCGCAATTTCACTCTGACTGTCACT

GACTTTTACCGCAGCTGGAACTGTGCTCCAGGGCCTTTCCATCTTTTCCCCCACACTCCCTTTGACCCTGTGTTGCC
CTCAGAGGCCAAATTCTTGGGCTCAGGGACTGGCTTTCGGCCTATTGGTGGTGGAGCTGGGGGTTCTGGCAAGGAAT
TTCAGGCAGCGCTTGGAGGCAATGTTCCCCGAGAGCAGTTCACGGTGGTGATGTTGACTTATGAGCGGGAGGAAGTG
CTTATGAACTCTTTAGAGAGGCTGAATGGCCTCCCTTACCTGAACAAGGTCGTGGTGGTGTGGAATTCTCCCAAGCT
GCCATCAGAGGACCTTCTGTGGCCTGACATTGGCGTCCCCATCATGGTGGTCCGTACTGAGAAGAACAGTTTGAACA
ACCGATTCTTACCCTGGAATGAAATTGAGACAGAGGCCATCCTGTCCATTGATGACGATGCTCACCTCCGCCATGAC
GAAATCATGTTTGGGTTCCGGGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCCTGGCCGTTACCACGCATG
GGACATCCCCCATCAGTCCTGGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGGTGCTGACAGGTGCTGCCT
TCTTTCACAAGGCCATCCGGGACATGGTGGATGAATACATCAACTGTGAGGACATTGCCATGAACTTCCTTGTCTCC
CACATCACTCGGAAGCCCCCCATCAAGGTGACCTCACGGTGGACATTCCGATGCCCAGGATGCCCTCAGGCCCTGTC
TCATGATGACTCCCACTTCCACGAGCGGCACAAGTGCATCAACTTCTTCGTGAAGGTGTACGGCTACATGCCCCTCC
TGTACACGCAGTTCAGGGTGGATTCTGTGCTCTTCAAGACACGCCTGCCCCATGACAAGACCAAGTGCTTCAAGTTC
ATCTAGGGGCAGCGCACGGTCTGGGGAAGAGGATGAGCAGAGGGAGGAAGATGGCTCCCAAGGTTCCTAGGCATTGC
AGGACCTTGGGCACATCTGCTGGTGGGTGGCCCAGAGCCTCTGCTGGAAGGGGCAGCAGGAGGAGTGGAAGGAAACC
GCTGCCTTTATCTTGAAGTCAGCCACACTGGGCCTGGAGCCCTGGGCGGAGTCCCCGGGGTTCCCCACACAGGGCAC
TGACTGATAGCTTACACTGAGGACTGTGGCGACTCTGCAGAGTCACTCACACCGTTCGTACGCCCAGGACAGCTGGT
TCGTGGTTTTTACATTCAATAACAACTATTATGATTATTTAAAAAGAGAAAGTTTCAGATTTGCCATTCAAGGCTTA
TTTATATATATGTGTGTGTATATAAATACATGCACACACTTGCATACATATATATTTTTGGCTGGGGGAGTGTGAGT
TTTGCCTTTCTAAGGGAGGGACCGCGCAGGCTCCTTTGTTCTGTATTCTGGCGGAGATGGGTCCTGGCCTTGTGTCA
CTGGCTTATCCTTAAAGATCATCTCCCATCCTCCCCAGCGCCATCTGTGTGCAGCAACCAGAAAGGGATGAACTTGG
CCCTCTTGCGGGCCTGGACAAGGTCTCTTCCTTACCCTTTCTGTTGCCAGTCAGCAACCTGTAACTCACATTCTCTT
CCCAGTGAATCCCTGGGAGCGCCTGACCCTGGTGGGCTGTTCAGCTTCCTGCTGCTGGGGCCAGCGATTTTTGAGGA
TTTATCTTTAGGCCAGGCTTGCCTCCGTACTTATCCCTGCTCTCCCATTTCTCTCTTGTTTGAGAGAGAATGAGGAA
GCAAAGAGTGAGAAAGAATAGGGGCTGAAGACGCCACTCCCAGATGGCTCTTTCTATCCTGCTCTTCTGTTGAAACA
CACGTGCTGTGGGCCTCAGGCGTTTCTGAAGTGCTCTTTCTTGGATTGGACAGGAGATCAGCAGCGTGCACATCTGC
TGTGGTCTGAAGTGGTTTGCAGGTCAGCCTCCTCTCCCTAGTGTAGAGCAAGCCAGTGTCCTTCGAGGAACCCACCC
GGCTGGCCGGGAAGTTTTACAGCAAGGCGCCTGCCTTGGGATAATTCCTTGGTGAAATTCACCTTCCCCCCGCCTCT
GTCTGGAGCCCCATCCTGTGTTATCTGTGGTTTTTGGACCCCTAATGTCAGCTTGGCTGTAGGACTCCCCGAGGTTT
GGTATGTGCTAGAACAATGGGAGGCTGTGATTTGCTGTGTAAGCTCACATCCAGCCTTGGAATCTAACGGGCATTCA
CAACCCGAGTTACCACTTTCCACTCCCTGCTTAGGATTCTGTTCCCTGGGCTGAAACTGAAATAAGCTAATTTTTTG
GGTCACGGTGGCAGTAGGGGAACCTAGGAGGGTGTGAGTGGCATTTGTCAGGGATTTAGCCCATGACGTGTTTCTTG
AACCCTACTTTCTGGAAGTGGAGTTGACTCTGGAAGTTTTCTAGCAACTGAACAAAAGCTCAGGTTTGTCCTGGTCA
TGCACATGCCTTAAGCCAGTTCCGTCTTCCCTAGACCTTGGCATCCTGTGCTTCTATTTCTTGGAATACGTTCTCCT
CTGACCTGCCTGTACCACGTGGGTCCTCTTCAAGTACTGTTTTGAAGCTGGGCTCTTTTGTGTAGCTCCCACCCACC
TGTAGGGCTAGCTCGGCTTAAGGGAACTCTCCCCATTGGCAAACCGGACCCGGCCGCCGCCAGGACTGTGTTTCCAA
AGGTTCCCCGCCCCCAACCCCAGCATCAGCCTGTAGCTCCCCTGCTGAGGCAGTGTGGTTATGTTCCCAGCAGTGGG
GGTCAGACGCCCTTCCTCAGAACTTTCTAGTTGCCCTCTACCTGACTCCTGACTTGTATTCCTTTTAGCAGTAGCCT
TCTTCCCTCGGGGAGCCAAAGAGTGTGGTGTGTGGCGCTATATTGTGGCTGCTATTTCATCTGGTTTCTTTTAATGT

GAGGAACTCACATACTGACTTCAGTGGGACTCGGTGAGCCGGGGCCGTCTGTGTGGTGGGACCCCCTTTAGCGGGAC
TCAGTGAGCTGGGGCCGTCTGTGTGGTGGAGCCAGGGCCTCTCCCTTTAGTGGAGCCAGGTTGTCGGGCCCCGAATG
TCACTGGTGGATCTAAGAAGGGCTGAGTGGTCTGACACCAAAACATGCCGCAGGGAGGGCTGTGGTGCCGGTGCTTC
CAACAAGGACAGCCCTCCTTGACCCTGAAAGGAACACTGGCTTGAAGGACTGCAGACAGGCTCTGAGGGGCACGCCC
TCCTCAGCGAGAGGCAGCAAGGTGGCCACAGTGTCACTGGTCAGGTGCTTCTCACCACGGGAAAGCCGCCGACCTGT
GACTCGCTTGAGATGGGAAAGCGGCGCCACAGACCCCGGGTCTCCTTGGCTGTCTGTGGGCCGCCCCTGGCCACCTT
GTCCTGGCTCGCAGGGTGCAGGAGCGCCTCGTTCTCTGGGTGGCCGGCTTGCTGCTCCGGTTTGGGCTGTCTTACCA
TAACACCGTCCCAGGGCTCTGCAGGCCACTGTGAGCGCTGGCTCCCTGGGCAGTGCTCCTCCGTGTGGACTGTGCCT
CAGGCCAGGGCTCACCAGCTGGGGTCCTGTCCGGAAGGATGGGATCTTTCTGGGAGCTGCGCCGGACAGAGTGGGGA
GCTCCTAGTTTGTGGGGGGAAGCTTTGATATCCATGCCACGTCCATCCACCCCACCCCTTTTCGTCACGAGCACAAT
GGTCTTACATTGGATTTTTGTAAAAAAATAAAAATAAATGGAGACTTTAACTCAAGCAGC

<210> SEQ ID NO 62

<211> Length : 3,466

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 62

>M79217_PEA_1_T10

TCCTGAGCTCAGGCAACCTGCCCGCCTCGGCCTCCCAGAGTGCTGGGATTACAGGCATGAGCCACGGTGCCCAGCCC
AGATGGGTTCTCACTTTATTGTTCAAGCTGGTCTGAAACTCCTGGCCTCGAGCAAGCCTCCCAAGTGCTGGGATTAC
AGGGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCCTGGCCGTTACCACGCATGGGACATCCCCCATCAGTC
CTGGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGGTGCTGACAGGTGCTGCCTTCTTTCACAAGTATTATG
CCTACCTGTATTCTTATGTGATGCCCCAGGCCATCCGGGACATGGTGGATGAATACATCAACTGTGAGGACATTGCC
ATGAACTTCCTTGTCTCCCACATCACTCGGAAGCCCCCCATCAAGGTGACCTCACGGTGGACATTCCGATGCCCAGG
ATGCCCTCAGGCCCTGTCTCATGATGACTCCCACTTCCACGAGCGGCACAAGTGCATCAACTTCTTCGTGAAGGTGT
ACGGCTACATGCCCCTCCTGTACACGCAGTTCAGGGTGGATTCTGTGCTCTTCAAGACACGCCTGCCCCATGACAAG
ACCAAGTGCTTCAAGTTCATCTAGGGGCAGCGCACGGTCTGGGGAAGAGGATGAGCAGAGGGAGGAAGATGGCTCCC
AAGGTTCCTAGGCATTGCAGGACCTTGGGCACATCTGCTGGTGGGTGGCCCAGAGCCTCTGCTGGAAGGGGCAGCAG
GAGGAGTGGAAGGAAACCGCTGCCTTTATCTTGAAGTCAGCCACACTGGGCCTGGAGCCCTGGGCGGAGTCCCCGGG
GTTCCCCACACAGGGCACTGACTGATAGCTTACACTGAGGACTGTGGCGACTCTGCAGAGTCACTCACACCGTTCGT
ACGCCCAGGACAGCTGGTTCGTGGTTTTTACATTCAATAACAACTATTATGATTATTTAAAAAGAGAAAGTTTCAGA
TTTGCCATTCAAGGCTTATTTATATATATGTGTGTGTATATAAATACATGCACACACTTGCATACATATATATTTTT
GGCTGGGGGAGTGTGAGTTTTGCCTTTCTAAGGGAGGGACCGCGCAGGCTCCTTTGTTCTGTATTCTGGCGGAGATG
GGTCCTGGCCTTGTGTCACTGGCTTATCCTTAAAGATCATCTCCCATCCTCCCCAGCGCCATCTGTGTGCAGCAACC
AGAAAGGGATGAACTTGGCCCTCTTGCGGGCCTGGACAAGGTCTCTTCCTTACCCTTTCTGTTGCCAGTCAGCAACC

TGTAACTCACATTCTCTTCCCAGTGAATCCCTGGGAGCGCCTGACCCTGGTGGGCTGTTCAGCTTCCTGCTGCTGGG
GCCAGCGATTTTTGAGGATTTATCTTTAGGCCAGGCTTGCCTCCGTACTTATCCCTGCTCTCCCATTTCTCTCTTGT
TTGAGAGAGAATGAGGAAGCAAAGAGTGAGAAAGAATAGGGGCTGAAGACGCCACTCCCAGATGGCTCTTTCTATCC
TGCTCTTCTGTTGAAACACACGTGCTGTGGGCCTCAGGCGTTTCTGAAGTGCTCTTTCTTGGATTGGACAGGAGATC
AGCAGCGTGCACATCTGCTGTGGTCTGAAGTGGTTTGCAGGTCAGCCTCCTCTCCCTAGTGTAGAGCAAGCCAGTGT
CCTTCGAGGAACCCACCCGGCTGGCCGGGAAGTTTTACAGCAAGGCGCCTGCCTTGGGATAATTCCTTGGTGAAATT
CACCTTCCCCCCGCCTCTGTCTGGAGCCCCATCCTGTGTTATCTGTGGTTTTTGGACCCCTAATGTCAGCTTGGCTG
TAGGACTCCCCGAGGTTTGGTATGTGCTAGAACAATGGGAGGCTGTGATTTGCTGTGTAAGCTCACATCCAGCCTTG
GAATCTAACGGGCATTCACAACCCGAGTTACCACTTTCCACTCCCTGCTTAGGATTCTGTTCCCTGGGCTGAAACTG
AAATAAGCTAATTTTTTGGGTCACGGTGGCAGTAGGGGAACCTAGGAGGGTGTGAGTGGCATTTGTCAGGGATTTAG
CCCATGACGTGTTTCTTGAACCCTACTTTCTGGAAGTGGAGTTGACTCTGGAAGTTTTCTAGCAACTGAACAAAAGC
TCAGGTTTGTCCTGGTCATGCACATGCCTTAAGCCAGTTCCGTCTTCCCTAGACCTTGGCATCCTGTGCTTCTATTT
CTTGGAATACGTTCTCCTCTGACCTGCCTGTACCACGTGGGTCCTCTTCAAGTACTGTTTTGAAGCTGGGCTCTTTT
GTGTAGCTCCCACCCACCTGTAGGGCTAGCTCGGCTTAAGGGAACTCTCCCCATTGGCAAACCGGACCCGGCCGCCG
CCAGGACTGTGTTTCCAAAGGTTCCCCGCCCCCAACCCCAGCATCAGCCTGTAGCTCCCCTGCTGAGGCAGTGTGGT
TATGTTCCCAGCAGTGGGGGTCAGACGCCCTTCCTCAGAACTTTCTAGTTGCCCTCTACCTGACTCCTGACTTGTAT
TCCTTTTAGCAGTAGCCTTCTTCCCTCGGGGAGCCAAAGAGTGTGGTGTGTGGCGCTATATTGTGGCTGCTATTTCA
TCTGGTTTCTTTTAATGTGAGGAACTCACATACTGACTTCAGTGGGACTCGGTGAGCCGGGGCCGTCTGTGTGGTGG
GACCCCCTTTAGCGGGACTCAGTGAGCTGGGGCCGTCTGTGTGGTGGAGCCAGGGCCTCTCCCTTTAGTGGAGCCAG
GTTGTCGGGCCCCGAATGTCACTGGTGGATCTAAGAAGGGCTGAGTGGTCTGACACCAAAACATGCCGCAGGGAGGG
CTGTGGTGCCGGTGCTTCCAACAAGGACAGCCCTCCTTGACCCTGAAAGGAACACTGGCTTGAAGGACTGCAGACAG
GCTCTGAGGGGCACGCCCTCCTCAGCGAGAGGCAGCAAGGTGGCCACAGTGTCACTGGTCAGGTGCTTCTCACCACG
GGAAAGCCGCCGACCTGTGACTCGCTTGAGATGGGAAAGCGGCGCCACAGACCCCGGGTCTCCTTGGCTGTCTGTGG
GCCGCCCCTGGCCACCTTGTCCTGGCTCGCAGGGTGCAGGAGCGCCTCGTTCTCTGGGTGGCCGGCTTGCTGCTCCG
GTTTGGGCTGTCTTACCATAACACCGTCCCAGGGCTCTGCAGGCCACTGTGAGCGCTGGCTCCCTGGGCAGTGCTCC
TCCGTGTGGACTGTGCCTCAGGCCAGGGCTCACCAGCTGGGGTCCTGTCCGGAAGGATGGGATCTTTCTGGGAGCTG
CGCCGGACAGAGTGGGGAGCTCCTAGTTTGTGGGGGGAAGCTTTGATATCCATGCCACGTCCATCCACCCCACCCCT
TTTCGTCACGAGCACAATGGTCTTACATTGGATTTTTGTAAAAAAATAAAAATAAATGGAGACTTTAACTCAAGCAG
C

<210> SEQ ID NO 63

<211> Length : 3,580

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 63

>M79217_PEA_1_T15

GGCGCCCTTCGGCAAGTTCCGCAGTCGCCTGTCGGAAATGGCTGCCGGCCGGCAGGGGGAGCGGCGGATCAGGCGCG
GCCTGGAAGGCGGGCGGCCGGCAGCCAGAACGGCTTCTGGGACGCCGACTTTCGCGCAGGCGGCGGCGGCGGCGGCG
GCGGGTCCCTGAGCTGGAAGCCGGAGAGCAAGCCCTGGAGGTTCACTCTTTCAAGAAGTCGTGTGCTGAGGTGTAAT
GCTACACAAGTCAGAGGAAGGAAGGGTCCTGAAACACATGGCCTGATTGTTGGCAAAGGCATCATAAGAAGCTGGCA
TTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATAGACACTATGCATATTTGTTGGTCAGCAAAACCAAGAA
ACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAGGGTGTTTTTCCTGGGTTTCATCATCAGGTACCTCCTC
CCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTGATAAAGATTAAGGACATGTTCTTTGGTCAACAGCCAG
AACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAGCTGCAGCTGAGGAAAATGAAATGTTCATTTTATTTGG
TGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTGTCAGTGAAACAGAGATCGTTTTGTGGAATAGCTAACC
CATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGGCAGGCTGCAGAGGACTCATGACAGGCTATACCATGCTGCG
GAATGGGGGCGCGGGGAACGGAGGTCAGACCTGCATGCTGCGCTGGTCCAACCGCATCCGCCTCACGTGGCTCAGCT
TCACGCTCTTTGTCATCCTGGTCTTCTTCCCGCTCATCGCCCACTATTACCTCACCACTCTGGATGAGGCTGATGAG
GCAGGCAAGCGGATTTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACGTGCTGGATCTGTGCCGCAT
CCGGGAGTCGGTGAGTGAAGAGCTCCTGCAGCTGGAGGCCAAGCGCCAAGAGCTGAACAGCGAGATCGCCAAGCTGA
ATCTGAAGATCGAAGCCTGTAAGAAGAGCATTGAGAACGCCAAGCAGGACCTGCTCCAGCTCAAGAATGTCATCAGC
CAGACCGAGCATTCCTACAAGGAGCTCATGGCCCAGAACCAGCCCAAGCTGTCCCTGCCCATCCGACTGCTCCCAGA
GAAGGACGATGCCGGCCTCCCTCCCCCGAAGGCCACTCGGGGCTGCCGGCTACACAACTGCTTTGATTATTCTCGTT
GCCCTCTCACCTCTGGCTTCCCGGTCTACGTCTATGACAGTGACCAGTTTGTCTTTGGCAGCTACCTGGATCCCTTG
GTCAAGCAGGCTTTTCAGGCGACAGCACGAGCTAACGTTTATGTTACAGAAAATGCAGACATCGCCTGCCTTTACGT
GATACTAGTGGGAGAGATGCAGGAGCCGGTGGTGCTGCGGCCTGCTGAGCTGGAGAAGCAGTTGTATTCCCTGCCAC
ACTGGCGGACGGATGGACACAACCATGTCATCATCAATCTGTCACGTAAGTCAGATACACAGAACCTTCTCTATAAC
GTCAGTACTGGCCGTGCCATGGTGGCCCAGTCCACCTTCTACACTGTCCAGTACAGACCTGGCTTTGACTTGGTCGT
ATCACCGCTGGTCCATGCCATGTCTGAGCCCAACTTCATGGAAATCCCACCACAGGTGCCGGTGAAGCGGAAATATC
TCTTCACCTTCCAGGGCGAGAAGATTGAGTCTCTGAGGTCTAGCCTTCAGGAGGCCCGCTCCTTCGAAGAGGAAATG
GAGGGCGACCCTCCCGCCGACTACGATGACCGGATCATTGCCACCCTGAAGGCGGTGCAGGACAGCAAGCTGGATCA
GGTCCTGGTGGAATTCACCTGCAAAAACCAGCCCAAACCCAGCCTGCCAACTGAGTGGGCACTGTGTGGAGAGCGGG
AGGACCGCTTGGAATTGCTGAAGCTCTCCACCTTCGCCCTCATCATTACCCCCGGGGACCCTCGCTTGGTTATTTCC
TCTGGGTGTGCAACACGGCTCTTCGAAGCCCTGGAAGTCGGTGCCGTCCCGGTGGTGCTGGGGGAGCAGGTCCAGCT
TCCCTACCAGGACATGCTGCAGTGGAACGAGGCGGCCCTGGTGGTGCCAAAGCCTCGTGTTACCGAGGTTCATTTCC
TGCTCAGAAGCCTCTCCGATAGTGACCTCCTGGCTATGAGGCGGCAAGGCCGCTTTCTCTGGGAGACTTACTTCTCC
ACTGCTGACAGTATTTTTAATACCGTGCTGGCTATGATTAGGACTCGCATCCAGATCCCAGCCGCTCCCATCCGGGA
AGAGGCGGCAGCTGAGATCCCCCACCGTTCAGGCAAGGCGGCTGGAACTGACCCCAACATGGCTGACAACGGGGACC
TGGACCTGGGGCCAGTGGAGACGGAGCCGCCCTACGCCTCACCCAGATACCTCCGCAATTTCACTCTGACTGTCACT
GACTTTTACCGCAGCTGGAACTGTGCTCCAGGGCCTTTCCATCTTTTCCCCCACACTCCCTTTGACCCTGTGTTGCC
CTCAGAGGCCAAATTCTTGGGCTCAGGGACTGGCTTTCGGCCTATTGGTGGTGGAGCTGGGGGTTCTGGCAAGGAAT
TTCAGGCAGCGCTTGGAGGCAATGTTCCCCGAGAGCAGTTCACGGTGGTGATGTTGACTTATGAGCGGGAGGAAGTG
CTTATGAACTCTTTAGAGAGGCTGAATGGCCTCCCTTACCTGAACAAGGTCGTGGTGGTGTGGAATTCTCCCAAGCT

GCCATCAGAGGACCTTCTGTGGCCTGACATTGGCGTCCCCATCATGGTGGTCCGTACTGAGAAGAACAGTTTGAACA
ACCGATTCTTACCCTGGAATGAAATTGAGACAGAGGCCATCCTGTCCATTGATGACGATGCTCACCTCCGCCATGAC
GAAATCATGTTTGGGTTCCGGGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCCTGGCCGTTACCACGCATG
GGACATCCCCCATCAGTCCTGGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGGTGCTGACAGGTGCTGCCT
TCTTTCACAAGGTAAGAAAAAGCTGGTAATAATGGCATCGACTTGGTGAGAGTTTCACCTTTGTGTGGTAGCGGAAT
GCTGCCCTCAGCTTAGCTCTCCTAACGCTTCTTACATGTTTCTTTTGTGCTAGAAGTCAGTTTTTTCTATTTTTACA
GACAATGATCAAGATGCTTAGAGCAACTCTGGGATAAAAAGTCAAGATGAGAGGGCTGCCTGTACAGTTGCACATAG
GCCATTTGGAAACCACTTTATCTTTCTGGGCGTTGGCTCTCCGTTTGTAAAACTGAGGGCACTGGGCTAAAGACACC
TCAAATACCTTCCAGTTTTAACACTGCCACCCTAGATATGGCCCAGCCATCAGAAGGTGACCTGGGCACTTTTCTGA
CTTAGATATACCATGCCTGTCCCGGGCCCCACGATGAG

<210> SEQ ID NO 64

<211> Length : 1,786

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 64

>M79217_PEA_1_T18

GGCGCCCTTCGGCAAGTTCCGCAGTCGCCTGTCGGAAATGGCTGCCGGCCGGCAGGGGGAGCGGCGGATCAGGCGCG
GCCTGGAAGGCGGGCGGCCGGCAGCCAGAACGGCTTCTGGGACGCCGACTTTCGCGCAGGCGGCGGCGGCGGCGGCG
GCGGGTCCCTGAGCTGGAAGCCGGAGAGCAAGCCCTGGAGGTTCACTCTTTCAAGAAGTCGTGTGCTGAGGTGTAAT
GCTACACAAGTCAGAGGAAGGAAGGGTCCTGAAACACATGGCCTGATTGTTGGCAAAGGCATCATAAGAAGCTGGCA
TTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATAGACACTATGCATATTTGTTGGTCAGCAAAACCAAGAA
ACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAGGGTGTTTTTCCTGGGTTTCATCATCAGGTACCTCCTC
CCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTGATAAAGATTAAGGACATGTTCTTTGGTCAACAGCCAG
AACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAGCTGCAGCTGAGGAAAATGAAATGTTCATTTTATTTGG
TGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTGTCAGTGAAACAGAGATCGTTTTGTGGAATAGCTAACC
CATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGCAGGCTGCAGAGGACTCATGACAGCCTGTAGCTCCCCTGC
TGAGGCAGTGTGGTTATGTTCCCAGCAGTGGGGGTCAGACGCCCTTCCTCAGAACTTTCTAGTTGCCCTCTACCTGA
CTCCTGACTTGTATTCCTTTTAGCAGTAGCCTTCTTCCCTCGGGGAGCCAAAGAGTGTGGTGTGTGGCGCTATATTG
TGGCTGCTATTTCATCTGGTTTCTTTTAATGTGAGGAACTCACATACTGACTTCAGTGGGACTCGGTGAGCCGGGGC
CGTCTGTGTGGTGGGACCCCCTTTAGCGGGACTCAGTGAGCTGGGGCCGTCTGTGTGGTGGAGCCAGGGCCTCTCCC
TTTAGTGGAGCCAGGTTGTCGGGCCCCGAATGTCACTGGTGGATCTAAGAAGGGCTGAGTGGTCTGACACCAAAACA
TGCCGCAGGGAGGGCTGTGGTGCCGGTGCTTCCAACAAGGACAGCCCTCCTTGACCCTGAAAGGAACACTGGCTTGA
AGGACTGCAGACAGGCTCTGAGGGGCACGCCCTCCTCAGCGAGAGGCAGCAAGGTGGCCACAGTGTCACTGGTCAGG
TGCTTCTCACCACGGGAAAGCCGCCGACCTGTGACTCGCTTGAGATGGGAAAGCGGCGCCACAGACCCCGGGTCTCC

TTGGCTGTCTGTGGGCCGCCCCTGGCCACCTTGTCCTGGCTCGCAGGGTGCAGGAGCGCCTCGTTCTCTGGGTGGCC
GGCTTGCTGCTCCGGTTTGGGCTGTCTTACCATAACACCGTCCCAGGGCTCTGCAGGCCACTGTGAGCGCTGGCTCC
CTGGGCAGTGCTCCTCCGTGTGGACTGTGCCTCAGGCCAGGGCTCACCAGCTGGGGTCCTGTCCGGAAGGATGGGAT
CTTTCTGGGAGCTGCGCCGGACAGAGTGGGGAGCTCCTAGTTTGTGGGGGGAAGCTTTGATATCCATGCCACGTCCA
TCCACCCCACCCCTTTTCGTCACGAGCACAATGGTCTTACATTGGATTTTTGTAAAAAAATAAAAATAAATGGAGAC
TTTAACTCAAGCAGC

<210> SEQ ID NO 65

<211> Length : 7,128

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 65

>M62096_PEA_1_T4

CGGCAGAGCGCGCTGGTGCTGATGCAGGATGGCTGAGCGCGCAGGAGCCCGGGAGGTCTGAGCCGGGCGAGGCTCGC
TCCCTGCGCATCGCCTCCTCCGCCCGCCGCGTGGTCGCGGGCAGGTGGGCCGGGGGGCGCTGGGCAGGGGCGGGGCA
GGGCCAGGGCAGGCCGGTCTGCAGCCGGAGGGGCCGGAGCGGAGAAGCTGCCCACCTTCCCGGGCTCGGAGCGGCCG
GGGCTGCTCAGCCGGCCGGGCTCGCGATGACCTGCTGAGAAGCGTCGTCGGAGGCTGCAGGAGGCGGCCTAGCTGTG
GGCGGTGCAGCTCGCGGCCTCCTCCCTCGTCGTTCCCGGCCCCGGCCCCCCACCCATCCCCGTGCCCCCTCCCTACC
GCCGGCCGAGATGGCGGATCCAGCCGAATGCAGCATCAAAGTGATGTGCCGGTTCCGGCCCCTCAACGAAGCGGAGA
TCCTCCGCGGGGACAAATTCATCCCCAAATTTAAAGGCGATGAGACCGTGGTGATCGGGCAAGGGAAGCCATATGTC
TTCGACAGAGTGCTACCTCCCAACACGACCCAAGAGCAGGTTTACAATGCATGTGCGAAGCAAATTGTCAAAGATGT
CCTTGAAGGTTATAACGGGACGATTTTTGCGTATGGGCAGACTTCATCAGGAAAAACCCACACCATGGAGGGGAAGC
TGCATGACCCCCAGCTCATGGGGATCATCCCACGAATTGCCCATGATATCTTTGACCATATCTACTCCATGGATGAG
AACCTGGAGTTTCACATAAAGGTTTCCTATTTTGAGATCTACTTGGACAAAATAAGGGACTTACTTGATGTATCCAA
GACCAACTTGGCTGTTCATGAAGATAAAAACAGAGTCCCGTATGTAAAGGGGTGCACTGAGCGGTTTGTGTCGAGCC
CTGAGGAAGTCATGGATGTAATAGATGAAGGCAAAGCAAACCGACACGTGGCTGTGACAAACATGAATGAACACAGC
TCTAGAAGTCACAGTATCTTCCTGATAAATATTAAACAAGAGAATGTAGAGACTGAAAAAAAACTCAGTGGGAAACT
TTATTTGGTTGATTTGGCTGGGAGCGAAAAGGTCAGCAAAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAA
ATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAAAAACACATGTGCCATAC
CGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACCATCGTCATTTGCTGTTC
TCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGAGCTAAGACCATCAAGAATACAG
TCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTTG
AAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATGATTTTCTAGCAGCACACGTGTTTGGAAA
GCTACTAGAATAATTGAATAATTCAGCACCTGAGGCTGGTGGATGATTCTTTGCAATTTGGCAGGAATGGGAGAGTC

GGGAGCAGTAGTTGGCAAGGTGGGGAGTAGCCATATGAAGTTTTATTTCGGGAATCCTCCAGGAGAAGCTGTGCCTG
AGGATGAACAGATCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATCATAGACAATATTGCT
CCTGTTGTTGCTGGCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCTCTACAGACAACTGGA
TGACAAGGATGATGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGATGAGC
TTTTAGCTTCCACAAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACACGTCTCCAGATTGAAAATGAGGCAGCC
AAGGATGAGGTGAAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAATTATGACCAGAAATCACAGGAAGTGGA
GGATAAGACCCGGGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAACGACTACATTGACAACCACACAGAGAG
AGCTGAGCCAGCTACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTGAGATCCTGAATTTGCTGTTGAAAGAT
CTGGGGGAGATAGGTGGAATTATTGGCACCAATGATGTGAAAACTTTGGCAGATGTGAATGGAGTCATTGAGGAGGA
GTTTACCATGGCCCGCCTGTACATCAGCAAGATGAAGTCAGAGGTCAAGTCCCTGGTGAACCGCAGCAAACAGCTCG
AGAGCGCCCAGATGGACTCCAACAGGAAGATGAATGCCAGCGAGCGGGAGCTGGCAGCCTGCCAGCTGCTCATCTCC
CAGCACGAAGCCAAGATCAAGTCTCTGACAGACTACATGCAGAACATGGAACAGAAGAGGAGGCAGCTAGAAGAGTC
CCAGGACTCGCTCAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAAAAATGCACGAAGTCAGCTTCCAGGATAAGG
AGAAGGAACATCTGACGCGGTTGCAGGATGCTGAAGAAATGAAGAAGGCGCTGGAGCAGCAGATGGAGAGCCACCGG
GAAGCTCACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAGCAGAAAATCATTGATGAGATTCGGGA
TTTGAATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAAGAGA
GAGAAATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCTGGAG
GAGACAGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAGTTAA
AAAAAGTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAAGCAGAAAATTTCCTTCTTGGAGAATA
ACCTGGAGCAGCTCACCAAAGTTCACAAGCAGCTGGTCCGGGACAACGCAGACCTGCGCTGTGAACTGCCCAAGCTG
GAGAAGCGGCTGCGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAGCGCGCTGAAGGAGGCCAAGGAGAACGCCAT
GCGGGACCGTAAGCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGGCCGTGCGGGCCAAGAACATGGCCAGAAGGG
CCCATTCAGCCCAGATCGCCAAGCCCATCCGCCCCGGACACTACCCGGCCTCATCTCCAACGGCCGTCCATGCCATT
CGAGGGGGAGGAGGCAGCTCTTCAAATTCCACTCACTACCAGAAATAAATACAAAATATGACTCCACGTAGCATGTC
AAGGACTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCCATTTTTTTTTTATACTTGCTTACTCC
AGCCATCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCTGTGTGTACAGATGTGTGCTCGGACTT
TTCTCTTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTACTGAGAACCATTACCACCGACTCGGC
CTCCGGGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCAGTGCACTGTCCCATCTGTACGCCCTA
ATGTGCCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATCATGGTCTGCCTCAACTGTCTGGTTTC
CTGTAAAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCTATAGGGTTCAAAGTTTTCCTTCTGAA
CACTTTTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCACATCTTTGGATCTGTAAAATATACCT
TTTAGTATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAAACAATCTGACAGTAGCAGTGTAGAAT
TTGTTCATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCTCCGAGCTGCTTGCTTTTGAACCTGCA
GCAACTAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTGAAAATGCTGCAAAGTTGGATAAGTGG
AAATGTGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCTTGCATTGGGAATGGCTGCTTTCTCTA
ACCATTTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAGCAGTTGAAAAAGGAAGCCTTCGGGAG
AAAGTGCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCCATGAATAAGTTATTATTTTAACCCAT
AATTGGCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCATCTTCTTAGGCACAATTAGAAAAAATC

CACATAGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACATGAATCTAAAGCTGAATCAACCCTTACTTC
CAGTTGTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGAAACTGACTGTGCTGCTATTTGTTTTG
ACAAATTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTCCTATCCTATCATGTTGTGTGCCCAAG
ATGAGTGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATATATGCCAGGGGAGACTTTGGGCTTTTC
TCATGACTGTGTGGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTG
TGTAAAGTGCTAAGAACTGTGCATTGACATCCAAACATTTCTTGTACAAAATTTCCCTAGCAAAGCAAACCTGCTTT
GACTTAATTTATTTGTTAAATGTTGCACTTTGTTTATGTATGTTTTGTTTTTGGTGGGGAATAAGGAGAGAGAGGAC
GACAAATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATTTTGCATTTGTTTAAATTTTTTTCATTCTTTAATTT
TGTTATCAGTGCCAGCCCAATATACCTGCTCTACCATTATTTGCGGTCTGATAAAAGGGTCCTTGTGGGGCAGGTTT
TGCAAAGCTTATCAGGTAATAACATATGCCACATAACCTTGTTGATATGTTTGCTTCTGATTTGGGAAGCTAAACAT
TGGTGTTTGAGAGGATTGCCAATTATTAATTGTCATTACCACTACTCTCCATTACTTTTTGTTTGGAAATTGAACAA
AGGTCAGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAAAATAGATTGTTTTAGATTCTTTCCAGGGTGATTTT
TCCCTGGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTCATGTTTCAAAGGGAAACATTCGCTTGTAGT
TCCATTTTACTTGATCTCTACAAGGGACTGACAACATTTGCTTTACTTTTATTCACAGAGAAAGTTGGCTTTGATGT
CTCTTAAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACCTAGCTTCAATCTTTATAGGACTTCTAATCT
AATTTTCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGAACGGATTATTCAAATGGATCCTTAAATATTGCTAT
GTATAATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTAGGCCACTTTCTAAAAAAGCCACATATGTGCAATTT
TCAGGTTTTTAGACTATTGCTCCCTGTACTTTAAATGTAAAAACCACACTTCTGAACAACTAAGCTCATGAATATGA
TTTTGGTTATATGCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCTCCTCTATGAATAATTTTATATTTCATGCTA
CTTCTTGAAAGTTTACTCTTTGATGCTCTAAGAGAACAGCCAGATGGTTTATATGAATAATCTTTATCTGCAGGATG
GTGGATTGGTAAATTAGGAGAATGTTGTTTGAGATATCAAGATTTATGTCTGGGAACTAAAATATATAATGCCAAAT
GTGTTTTTGTCAATTACTAGAGAATTCTGTGCAAACATATCATCTCTTCAAATGCTGCACACTTTGCTTTTGTTAAA
CAGCAGGTAGTAGACAGAACAATAACAGTTTCGCGTTAAGACTTTTAAAGGAAATAGAATCGTGATTAAGAAATCAG
AATTTATAGATATATTGGGATAAATGAAGAAATAAAAATGTTTGTCTAGAATGTAGCATCTAGTGACTTTTTAAAGC
CCTAACGTTTACATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAGCAAATAAAAGTTCTTAACAATCCCCTCTTT
CGAAGTGCATTTTTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGCTATATGACATACTGGGAAAGGCAGAGGGTG
GAGGGAAGATTTCACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCTAAAGAAACCATCATAATCTAAAATTGCTTC
ATTTAACACTAACAATTTAGACTTTTTAAACCAAGCATTGAATAATGGCTGGATAACTGCCGAAGTAAGCGCCGCTC
CATGAAGTCTGCTTACTTATTTAAAAATTGTGTATCAGTTTTAAATACTGTTCATTGTGTGCAGATATAAGGGGAAT
AGGGCATTCTGTAGAATTATACATGTCTAGTTTGTAAAGTGTGTCCTGTGTACTGCAGATGTGTGTTCTCTGGGCTT
TATGTATCTGTACAGTAGCTTTCACATTAAAAAAATTGTGGACAAACTTGTCCGGGGGGGTTTGAGGGGAGAATGGTG
GTTTATATCAATAACGATGCTGTACTATAGTCCATGTAACAAAAGATCTGGAAGTCACCCTCCTCTGGCCCACGGAA
AATTTTGGTAATCTTCTAGGTTCTAAAATGAAGATGTATGGGTACTCTGGCAGACTGCATGTTGTATAATTTGAAAA
ATACTAAAAGTGGAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 66

<211> Length : 7,004

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 66

>M62096_PEA_1_T5

CGGCAGAGCGCGCTGGTGCTGATGCAGGATGGCTGAGCGCGCAGGAGCCCGGGAGGTCTGAGCCGGGCGAGGCTCGC
TCCCTGCGCATCGCCTCCTCCGCCCGCCGCGTGGTCGCGGGCAGGTGGGCCGGGGGGCGCTGGGCAGGGGCGGGGCA
GGGCCAGGGCAGGCCGGTCTGCAGCCGGAGGGGCCGGAGCGGAGAAGCTGCCCACCTTCCCGGGCTCGGAGCGGCCG
GGGCTGCTCAGCCGGCCGGGCTCGCGATGACCTGCTGAGAAGCGTCGTCGGAGGCTGCAGGAGGCGGCCTAGCTGTG
GGCGGTGCAGCTCGCGGCCTCCTCCCTCGTCGTTCCCGGCCCCGGCCCCCCTCCCTACCGCCGGCCGAGATGGCGGA
TCCAGCCGAATGCAGCATCAAAGTGATGTGCCGGTTCCGGCCCCTCAACGAAGCGGAGATCCTCCGCGGGGACAAAT
TCATCCCCAAATTTAAAGGCGATGAGACCGTGGTGATCGGGCAAGGGAAGCCATATGTCTTCGACAGAGTGCTACCT
CCCAACACGACCCAAGAGCAGGTTTACAATGCATGTGCGAAGCAAATTGTCAAAGATGTCCTTGAAGGTTATAACGG
GACGATTTTTGCGTATGGGCAGACTTCATCAGGAAAAACCCACACCATGGAGGGGAAGCTGCATGACCCCCAGCTCA
TGGGGATCATCCCACGAATTGCCCATGATATCTTTGACCATATCTACTCCATGGATGAGAACCTGGAGTTTCACATA
AAGGTTTCCTATTTTGAGATCTACTTGGACAAAATAAGGGACTTACTTGATGTATCCAAGACCAACTTGGCTGTTCA
TGAAGATAAAAACAGAGTCCCGTATGTAAAGGGGTGCACTGAGCGGTTTGTGTCGAGCCCTGAGGAAGTCATGGATG
TAATAGATGAAGGCAAAGCAAACCGACACGTGGCTGTGACAAACATGAATGAACACAGCTCTAGAAGTCACAGTATC
TTCCTGATAAATATTAAACAAGAGAATGTAGAGACTGAAAAAAAACTCAGTGGGAAACTTTATTTGGTTGATTTGGC
TGGGAGCGAAAAGGTCAGCAAAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAAATATCAATAAGTCTTTGT
CTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAAAAACACATGTGCCATACCGGGACAGCAAGATGACT
CGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACCATCGTCATTTGCTGTTCTCCTTCTGTCTTCAATGA
GGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGGGTATGAAATCAAGGCTTAGGTGCAAAGCCATTGGATACCA
TACCTGAGACCACACAGCCAAGCTAAGACCATCAAGAATACAGTCTCTGTGAACCTAGAACTGACAGCAGAAGAATG
GAAGAAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTTGAAGAATGTTATCCAGCATCTGGAGATGGAGCTAA
ACAGGTGGAGGAATGGAGAAGCTGTGCCTGAGGATGAACAGATCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGT
GATAACACCCCCATCATAGACAATATTGCTCCTGTTGTTGCTGGCATCTCTACAGAGGAGAAAGAGAAGTACGATGA
GGAGATCTCCAGTCTCTACAGACAACTGGATGACAAGGATGATGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGC
TGAAGCAACAGATGTTGGATCAGGATGAGCTTTTAGCTTCCACAAGAAGAGACTATGAGAAGATACAGGAGGAGCTG
ACACGTCTCCAGATTGAAAATGAGGCAGCCAAGGATGAGGTGAAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGT
CAATTATGACCAGAAATCACAGGAAGTGGAGGATAAGACCCGGGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGA
AAACGACTACATTGACAACCACACAGAGAGAGCTGAGCCAGCTACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCA
ACTGAGATCCTGAATTTGCTGTTGAAAGATCTGGGGGGAGATAGGTGGAATTATTGGCACCAATGATGTGAAAACTTT
GGCAGATGTGAATGGAGTCATTGAGGAGGAGTTTACCATGGCCCGCCTGTACATCAGCAAGATGAAGTCAGAGGTCA
AGTCCCTGGTGAACCGCAGCAAACAGCTCGAGAGCGCCCAGATGGACTCCAACAGGAAGATGAATGCCAGCGAGCGG
GAGCTGGCAGCCTGCCAGCTGCTCATCTCCCAGCACGAAGCCAAGATCAAGTCTCTGACAGACTACATGCAGAACAT
GGAACAGAAGAGGAGGCAGCTAGAAGAGTCCCAGGACTCGCTCAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAA

AAATGCACGAAGTCAGCTTCCAGGATAAGGAGAAGGAACATCTGACGCGGTTGCAGGATGCTGAAGAAATGAAGAAG
GCGCTGGAGCAGCAGATGGAGAGCCACCGGGAAGCTCACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGA
GAAGCAGAAAATCATTGATGAGATTCGGGATTTGAATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATT
ATAACAAGCTGAAAATAGAGGACCAAGAGAGAGAAATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAA
CAAGCCAGAGAAGACCTCAAAGGGCTGGAGGAGACAGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACT
CTTTGTCCAGGATCTGACCACCCGAGTTAAAAAAAGTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCC
AGAAGCAGAAAATTTCCTTCTTGGAGAATAACCTGGAGCAGCTCACCAAAGTTCACAAGCAGCTGGTCCGGGACAAC
GCAGACCTGCGCTGTGAACTGCCCAAGCTGGAGAAGCGGCTGCGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAG
CGCGCTGAAGGAGGCCAAGGAGAACGCCATGCGGGACCGTAAGCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGG
CCGTGCGGGCCAAGAACATGGCCAGAAGGGCCCATTCAGCCCAGATCGCCAAGCCCATCCGCCCCGGACACTACCCG
GCCTCATCTCCAACGGCCGTCCATGCCATTCGAGGGGGAGGAGGCAGCTCTTCAAATTCCACTCACTACCAGAAATA
AATACAAAATATGACTCCACGTAGCATGTCAAGGACTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGT
TTCCATTTTTTTTTTATACTTGCTTACTCCAGCCATCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGT
TTCTGTGTGTACAGATGTGTGCTCGGACTTTTCTCTTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTT
ACTACTGAGAACCATTACCACCGACTCGGCCTCCGGGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTG
GGCAGTGCACTGTCCCATCTGTACGCCCTAATGTGCCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGAT
GATCATGGTCTGCCTCAACTGTCTGGTTTCCTGTAAAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGC
TGCTATAGGGTTCAAAGTTTTCCTTCTGAACACTTTTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTG
GTCACATCTTTGGATCTGTAAAATATACCTTTTAGTATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAG
AAAAACAATCTGACAGTAGCAGTGTAGAATTTGTTCATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTC
ACCTCCGAGCTGCTTGCTTTTGAACCTGCAGCAACTAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGT
GCTGAAAATGCTGCAAAGTTGGATAAGTGGAAATGTGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGA
AGCTTGCATTGGGAATGGCTGCTTTCTCTAACCATTTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATT
CCAGCAGTTGAAAAAGGAAGCCTTCGGGAGAAAGTGCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGT
TGCCATGAATAAGTTATTATTTTAACCCATAATTGGCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTT
TCATCTTCTTAGGCACAATTAGAAAAAATCCACATAGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACA
TGAATCTAAAGCTGAATCAACCCTTACTTCCAGTTGTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCA
GGGAAACTGACTGTGCTGCTATTTGTTTTGACAAATTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAA
ACTCCTATCCTATCATGTTGTGTGCCCAAGATGAGTGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAA
ATATATGCCAGGGGAGACTTTGGGCTTTTCTCATGACTGTGTGGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTAAAGTGCTAAGAACTGTGCATTGACATCCAAACATTTCTTGTAC
AAAATTTCCCTAGCAAAGCAAACCTGCTTTGACTTAATTTATTTGTTAAATGTTGCACTTTGTTTATGTATGTTTTG
TTTTTGGTGGGGAATAAGGAGAGAGAGGACGACAAATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATTTTGCA
TTTGTTTAAATTTTTTTCATTCTTTAATTTTGTTATCAGTGCCAGCCCAATATACCTGCTCTACCATTATTTGCGGT
CTGATAAAAGGGTCCTTGTGGGGCAGGTTTTGCAAAGCTTATCAGGTAATAACATATGCCACATAACCTTGTTGATA
TGTTTGCTTCTGATTTGGGAAGCTAAACATTGGTGTTTGAGAGGATTGCCAATTATTAATTGTCATTACCACTACTC
TCCATTACTTTTTGTTTGGAAATTGAACAAAGGTCAGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAAAATAG
ATTGTTTTAGATTCTTTCCAGGGTGATTTTTCCCTGGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTC

ATGTTTCAAAGGGAAACATTCGCTTGTAGTTCCATTTTACTTGATCTCTACAAGGGACTGACAACATTTGCTTTACT
TTTATTCACAGAGAAAGTTGGCTTTGATGTCTCTTAAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACC
TAGCTTCAATCTTTATAGGACTTCTAATCTAATTTTCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGAACGGA
TTATTCAAATGGATCCTTAAATATTGCTATGTATAATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTAGGCCA
CTTTCTAAAAAAGCCACATATGTGCAATTTTCAGGTTTTTAGACTATTGCTCCCTGTACTTTAAATGTAAAAACCAC
ACTTCTGAACAACTAAGCTCATGAATATGATTTTGGTTATATGCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCT
CCTCTATGAATAATTTTATATTTCATGCTACTTCTTGAAAGTTTACTCTTTGATGCTCTAAGAGAACAGCCAGATGG
TTTATATGAATAATCTTTATCTGCAGGATGGTGGATTGGTAAATTAGGAGAATGTTGTTTGAGATATCAAGATTTAT
GTCTGGGAACTAAAATATATAATGCCAAATGTGTTTTTGTCAATTACTAGAGAATTCTGTGCAAACATATCATCTCT
TCAAATGCTGCACACTTTGCTTTTGTTAAACAGCAGGTAGTAGACAGAACAATAACAGTTTCGCGTTAAGACTTTTA
AAGGAAATAGAATCGTGATTAAGAAATCAGAATTTATAGATATATTGGGATAAATGAAGAAATAAAAATGTTTGTCT
AGAATGTAGCATCTAGTGACTTTTTAAAGCCCTAACGTTTACATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAG
CAAATAAAAGTTCTTAACAATCCCCTCTTTCGAAGTGCATTTTTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGC
TATATGACATACTGGGAAAGGCAGAGGGTGGAGGGAAGATTTCACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCT
AAAGAAACCATCATAATCTAAAATTGCTTCATTTAACACTAACAATTTAGACTTTTTAAACCAAGCATTGAATAATG
GCTGGATAACTGCCGAAGTAAGCGCCGCTCCATGAAGTCTGCTTACTTATTTAAAAATTGTGTATCAGTTTTAAATA
CTGTTCATTGTGTGCAGATATAAGGGGAATAGGGCATTCTGTAGAATTATACATGTCTAGTTTGTAAAGTGTGTCCT
GTGTACTGCAGATGTGTGTTCTCTGGGCTTTATGTATCTGTACAGTAGCTTTCACATTAAAAAAATTGTGGACAAAC
TTGTCCGGGGGGTTTGAGGGGAGAATGGTGGTTTATATCAATAACGATGCTGTACTATAGTCCATGTAACAAAAGAT
CTGGAAGTCACCCTCCTCTGGCCCACGGAAAATTTTGGTAATCTTCTAGGTTCTAAAATGAAGATGTATGGGTACTC
TGGCAGACTGCATGTTGTATAATTTGAAAAATACTAAAAGTGGAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 67

<211> Length : 5,977

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 67

>M62096_PEA_1_T6

GCTGATTGTCCCCATGAAGGCCAGCCTTGAAGCTTGGTCAGTCTCCCTAACTGTATGATTGATCCCCACTTATTGCA
CTACATCACTGAGTTCCCGTATGCCAAGTTATGGCCACTTACATCCACGTCAGCAAAACTGGTGCCGAGGGAGCTGT
TCTTGACGAAGCTAAAAATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAA
AAACACATGTGCCATACCGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACC
ATCGTCATTTGCTGTTCTCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGAGCTAA
GACCATCAAGAATACAGTCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAGAGAAAG
AGAAAAACAAGACTTTGAAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATGGAGAAGCTGTG

CCTGAGGATGAACAGATCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATCATAGACAATAT
TGCTCCTGTTGTTGCTGGCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCTCTACAGACAAC
TGGATGACAAGGATGATGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGAT
GAGCTTTTAGCTTCCACAAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACACGTCTCCAGATTGAAAATGAGGC
AGCCAAGGATGAGGTGAAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAATTATGACCAGAAATCACAGGAAG
TGGAGGATAAGACCCGGGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAACGACTACATTGACAACCACACAG
AGAGAGCTGAGCCAGCTACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTGAGATCCTGAATTTGCTGTTGAA
AGATCTGGGGGAGATAGGTGGAATTATTGGCACCAATGATGTGAAAACTTTGGCAGATGTGAATGGAGTCATTGAGG
AGGAGTTTACCATGGCCCGCCTGTACATCAGCAAGATGAAGTCAGAGGTCAAGTCCCTGGTGAACCGCAGCAAACAG
CTCGAGAGCGCCCAGATGGACTCCAACAGGAAGATGAATGCCAGCGAGCGGGAGCTGGCAGCCTGCCAGCTGCTCAT
CTCCCAGCACGAAGCCAAGATCAAGTCTCTGACAGACTACATGCAGAACATGGAACAGAAGAGGAGGCAGCTAGAAG
AGTCCCAGGACTCGCTCAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAAAAATGCACGAAGTCAGCTTCCAGGAT
AAGGAGAAGGAACATCTGACGCGGTTGCAGGATGCTGAAGAAATGAAGAAGGCGCTGGAGCAGCAGATGGAGAGCCA
CCGGGAAGCTCACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAGCAGAAAATCATTGATGAGATTC
GGGATTTGAATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAA
GAGAGAGAAATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCT
GGAGGAGACAGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAG
TTAAAAAAAGTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAAGCAGAAAATTTCCTTCTTGGAG
AATAACCTGGAGCAGCTCACCAAAGTTCACAAGCAGCTGGTCCGGGACAACGCAGACCTGCGCTGTGAACTGCCCAA
GCTGGAGAAGCGGCTGCGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAGCGCGCTGAAGGAGGCCAAGGAGAACG
CCATGCGGGACCGTAAGCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGGCCGTGCGGGCCAAGAACATGGCCAGA
AGGGCCCATTCAGCCCAGATCGCCAAGCCCATCCGCCCCGGACACTACCCGGCCTCATCTCCAACGGCCGTCCATGC
CATTCGAGGGGGAGGAGGCAGCTCTTCAAATTCCACTCACTACCAGAAATAAATACAAAATATGACTCCACGTAGCA
TGTCAAGGACTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCCATTTTTTTTTTATACTTGCTTA
CTCCAGCCATCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCTGTGTGTACAGATGTGTGCTCGG
ACTTTTCTCTTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTACTGAGAACCATTACCACCGACT
CGGCCTCCGGGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCAGTGCACTGTCCCATCTGTACGC
CCTAATGTGCCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATCATGGTCTGCCTCAACTGTCTGG
TTTCCTGTAAAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCTATAGGGTTCAAAGTTTTCCTTC
TGAACACTTTTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCACATCTTTGGATCTGTAAAATAT
ACCTTTTAGTATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAAACAATCTGACAGTAGCAGTGTA
GAATTTGTTCATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCTCCGAGCTGCTTGCTTTTGAACC
TGCAGCAACTAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTGAAAATGCTGCAAAGTTGGATAA
GTGGAAATGTGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCTTGCATTGGGAATGGCTGCTTTC
TCTAACCATTTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAGCAGTTGAAAAAGGAAGCCTTCG
GGAGAAAGTGCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCCATGAATAAGTTATTATTTTAAC
CCATAATTGGCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCATCTTCTTAGGCACAATTAGAAAA
AATCCACATAGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACATGAATCTAAAGCTGAATCAACCCTTA

CTTCCAGTTGTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGAAACTGACTGTGCTGCTATTTGT
TTTGACAAATTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTCCTATCCTATCATGTTGTGTGCC
CAAGATGAGTGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATATATGCCAGGGGAGACTTTGGGCT
TTTCTCATGACTGTGTGGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTA
TGTGTGTAAAGTGCTAAGAACTGTGCATTGACATCCAAACATTTCTTGTACAAAATTTCCCTAGCAAAGCAAACCTG
CTTTGACTTAATTTATTTGTTAAATGTTGCACTTTGTTTATGTATGTTTTGTTTTTGGTGGGGAATAAGGAGAGAGA
GGACGACAAATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATTTTGCATTTGTTTAAATTTTTTTCATTCTTTA
ATTTTGTTATCAGTGCCAGCCCAATATACCTGCTCTACCATTATTTGCGGTCTGATAAAAGGGTCCTTGTGGGGCAG
GTTTTGCAAAGCTTATCAGGTAATAACATATGCCACATAACCTTGTTGATATGTTTGCTTCTGATTTGGGAAGCTAA
ACATTGGTGTTTGAGAGGATTGCCAATTATTAATTGTCATTACCACTACTCTCCATTACTTTTTGTTTGGAAATTGA
ACAAAGGTCAGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAAAATAGATTGTTTTAGATTCTTTCCAGGGTGA
TTTTTCCCTGGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTCATGTTTCAAAGGGAAACATTCGCTTG
TAGTTCCATTTTACTTGATCTCTACAAGGGACTGACAACATTTGCTTTACTTTTATTCACAGAGAAAGTTGGCTTTG
ATGTCTCTTAAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACCTAGCTTCAATCTTTATAGGACTTCTA
ATCTAATTTTCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGAACGGATTATTCAAATGGATCCTTAAATATTG
CTATGTATAATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTAGGCCACTTTCTAAAAAAGCCACATATGTGCA
ATTTTCAGGTTTTTAGACTATTGCTCCCTGTACTTTAAATGTAAAAACCACACTTCTGAACAACTAAGCTCATGAAT
ATGATTTTGGTTATATGCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCTCCTCTATGAATAATTTTATATTTCAT
GCTACTTCTTGAAAGTTTACTCTTTGATGCTCTAAGAGAACAGCCAGATGGTTTATATGAATAATCTTTATCTGCAG
GATGGTGGATTGGTAAATTAGGAGAATGTTGTTTGAGATATCAAGATTTATGTCTGGGAACTAAAATATATAATGCC
AAATGTGTTTTTGTCAATTACTAGAGAATTCTGTGCAAACATATCATCTCTTCAAATGCTGCACACTTTGCTTTTGT
TAAACAGCAGGTAGTAGACAGAACAATAACAGTTTCGCGTTAAGACTTTTAAAGGAAATAGAATCGTGATTAAGAAA
TCAGAATTTATAGATATATTGGGATAAATGAAGAAATAAAAATGTTTGTCTAGAATGTAGCATCTAGTGACTTTTTA
AAGCCCTAACGTTTACATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAGCAAATAAAAGTTCTTAACAATCCCCT
CTTTCGAAGTGCATTTTTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGCTATATGACATACTGGGAAAGGCAGAG
GGTGGAGGGAAGATTTCACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCTAAAGAAACCATCATAATCTAAAATTG
CTTCATTTAACACTAACAATTTAGACTTTTTAAACCAAGCATTGAATAATGGCTGGATAACTGCCGAAGTAAGCGCC
GCTCCATGAAGTCTGCTTACTTATTTAAAAATTGTGTATCAGTTTTAAATACTGTTCATTGTGTGCAGATATAAGGG
GAATAGGGCATTCTGTAGAATTATACATGTCTAGTTTGTAAAGTGTGTCCTGTGTACTGCAGATGTGTGTTCTCTGG
GCTTTATGTATCTGTACAGTAGCTTTCACATTAAAAAAATTGTGGACAAACTTGTCCGGGGGGTTTGAGGGGAGAAT
GGTGGTTTATATCAATAACGATGCTGTACTATAGTCCATGTAACAAAAGATCTGGAAGTCACCCTCCTCTGGCCCAC
GGAAAATTTTGGTAATCTTCTAGGTTCTAAAATGAAGATGTATGGGTACTCTGGCAGACTGCATGTTGTATAATTTG
AAAAATACTAAAAGTGGAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 68

<211> Length : 5,999

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 68

>M62096_PEA_1_T7

GGTTTGACTTAGCCTGTGGCAGAGGCAGCTGCACACATGTGGGAGGGATATGAACTTCTGAGAAAAGAGAAAATCCC
ATGTTGTACCCGACTCTAATAGGCCAGGAACGAGCTTTGCCATTGGAATGTGGCAGTCCTGCCTCTGCAGGTCAGCA
AAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAAATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATC
TCTGCTTTGGCAGAAGGGACAAAAACACATGTGCCATACCGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTT
GGGTGGGAACTGCAGAACCACCATCGTCATTTGCTGTTCTCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACAC
TGATGTTCGGACAGAGAGCTAAGACCATCAAGAATACAGTCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAG
AAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTTGAAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAG
GTGGAGGAATGGAGAAGCTGTGCCTGAGGATGAACAGATCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATA
ACACCCCCATCATAGACAATATTGCTCCTGTTGTTGCTGGCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAG
ATCTCCAGTCTCTACAGACAACTGGATGACAAGGATGATGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAA
GCAACAGATGTTGGATCAGGATGAGCTTTTAGCTTCCACAAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACAC
GTCTCCAGATTGAAAATGAGGCAGCCAAGGATGAGGTGAAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAAT
TATGACCAGAAATCACAGGAAGTGGAGGATAAGACCCGGGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAAC
GACTACATTGACAACCACACAGAGAGAGCTGAGCCAGCTACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTG
AGATCCTGAATTTGCTGTTGAAAGATCTGGGGGAGATAGGTGGAATTATTGGCACCAATGATGTGAAAACTTTGGCA
GATGTGAATGGAGTCATTGAGGAGGAGTTTACCATGGCCCGCCTGTACATCAGCAAGATGAAGTCAGAGGTCAAGTC
CCTGGTGAACCGCAGCAAACAGCTCGAGAGCGCCCAGATGGACTCCAACAGGAAGATGAATGCCAGCGAGCGGGAGC
TGGCAGCCTGCCAGCTGCTCATCTCCCAGCACGAAGCCAAGATCAAGTCTCTGACAGACTACATGCAGAACATGGAA
CAGAAGAGGAGGCAGCTAGAAGAGTCCCAGGACTCGCTCAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAAAAAT
GCACGAAGTCAGCTTCCAGGATAAGGAGAAGGAACATCTGACGCGGTTGCAGGATGCTGAAGAAATGAAGAAGGCGC
TGGAGCAGCAGATGGAGAGCCACCGGGAAGCTCACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAG
CAGAAAATCATTGATGAGATTCGGGATTTGAATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAA
CAAGCTGAAAATAGAGGACCAAGAGAGAGAAATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAG
CCAGAGAAGACCTCAAAGGGCTGGAGGAGACAGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTT
GTCCAGGATCTGACCACCCGAGTTAAAAAAAGTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAA
GCAGAAAATTTCCTTCTTGGAGAATAACCTGGAGCAGCTCACCAAAGTTCACAAGCAGCTGGTCCGGGACAACGCAG
ACCTGCGCTGTGAACTGCCCAAGCTGGAGAAGCGGCTGCGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAGCGCG
CTGAAGGAGGCCAAGGAGAACGCCATGCGGGACCGTAAGCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGGCCGT
GCGGGCCAAGAACATGGCCAGAAGGGCCCATTCAGCCCAGATCGCCAAGCCCATCCGCCCCGGACACTACCCGGCCT
CATCTCCAACGGCCGTCCATGCCATTCGAGGGGGAGGAGGCAGCTCTTCAAATTCCACTCACTACCAGAAATAAATA
CAAAATATGACTCCACGTAGCATGTCAAGGACTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCC
ATTTTTTTTTTATACTTGCTTACTCCAGCCATCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCT
GTGTGTACAGATGTGTGCTCGGACTTTTCTCTTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTA

CTGAGAACCATTACCACCGACTCGGCCTCCGGGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCA
GTGCACTGTCCCATCTGTACGCCCTAATGTGCCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATC
ATGGTCTGCCTCAACTGTCTGGTTTCCTGTAAAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCT
ATAGGGTTCAAAGTTTTCCTTCTGAACACTTTTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCA
CATCTTTGGATCTGTAAAATATACCTTTTAGTATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAA
ACAATCTGACAGTAGCAGTGTAGAATTTGTTCATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCT
CCGAGCTGCTTGCTTTTGAACCTGCAGCAACTAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTG
AAAATGCTGCAAAGTTGGATAAGTGGAAATGTGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCT
TGCATTGGGAATGGCTGCTTTCTCTAACCATTTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAG
CAGTTGAAAAAGGAAGCCTTCGGGAGAAAGTGCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCC
ATGAATAAGTTATTATTTTAACCCATAATTGGCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCAT
CTTCTTAGGCACAATTAGAAAAAATCCACATAGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACATGAA
TCTAAAGCTGAATCAACCCTTACTTCCAGTTGTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGA
AACTGACTGTGCTGCTATTTGTTTTGACAAATTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTC
CTATCCTATCATGTTGTGTGCCCAAGATGAGTGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATAT
ATGCCAGGGGAGACTTTGGGCTTTTCTCATGACTGTGTGGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTATGTGTGTAAAGTGCTAAGAACTGTGCATTGACATCCAAACATTTCTTGTACAAAA
TTTCCCTAGCAAAGCAAACCTGCTTTGACTTAATTTATTTGTTAAATGTTGCACTTTGTTTATGTATGTTTTGTTTT
TGGTGGGGAATAAGGAGAGAGAGGACGACAAATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATTTTGCATTTG
TTTAAATTTTTTTCATTCTTTAATTTTGTTATCAGTGCCAGCCCAATATACCTGCTCTACCATTATTTGCGGTCTGA
TAAAAGGGTCCTTGTGGGGCAGGTTTTGCAAAGCTTATCAGGTAATAACATATGCCACATAACCTTGTTGATATGTT
TGCTTCTGATTTGGGAAGCTAAACATTGGTGTTTGAGAGGATTGCCAATTATTAATTGTCATTACCACTACTCTCCA
TTACTTTTTGTTTGGAAATTGAACAAAGGTCAGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAAAATAGATTG
TTTTAGATTCTTTCCAGGGTGATTTTTCCCTGGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTCATGT
TTCAAAGGGAAACATTCGCTTGTAGTTCCATTTTACTTGATCTCTACAAGGGACTGACAACATTTGCTTTACTTTTA
TTCACAGAGAAAGTTGGCTTTGATGTCTCTTAAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACCTAGC
TTCAATCTTTATAGGACTTCTAATCTAATTTTCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGAACGGATTAT
TCAAATGGATCCTTAAATATTGCTATGTATAATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTAGGCCACTTT
CTAAAAAAGCCACATATGTGCAATTTTCAGGTTTTTAGACTATTGCTCCCTGTACTTTAAATGTAAAAACCACACTT
CTGAACAACTAAGCTCATGAATATGATTTTGGTTATATGCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCTCCTC
TATGAATAATTTTATATTTCATGCTACTTCTTGAAAGTTTACTCTTTGATGCTCTAAGAGAACAGCCAGATGGTTTA
TATGAATAATCTTTATCTGCAGGATGGTGGATTGGTAAATTAGGAGAATGTTGTTTGAGATATCAAGATTTATGTCT
GGGAACTAAAATATATAATGCCAAATGTGTTTTTGTCAATTACTAGAGAATTCTGTGCAAACATATCATCTCTTCAA
ATGCTGCACACTTTGCTTTTGTTAAACAGCAGGTAGTAGACAGAACAATAACAGTTTCGCGTTAAGACTTTTAAAGG
AAATAGAATCGTGATTAAGAAATCAGAATTTATAGATATATTGGGATAAATGAAGAAATAAAAATGTTTGTCTAGAA
TGTAGCATCTAGTGACTTTTTAAAGCCCTAACGTTTACATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAGCAAA
TAAAAGTTCTTAACAATCCCCTCTTTCGAAGTGCATTTTTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGCTATA
TGACATACTGGGAAAGGCAGAGGGTGGAGGGAAGATTTCACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCTAAAG

AAACCATCATAATCTAAAATTGCTTCATTTAACACTAACAATTTAGACTTTTTAAACCAAGCATTGAATAATGGCTG
GATAACTGCCGAAGTAAGCGCCGCTCCATGAAGTCTGCTTACTTATTTAAAAATTGTGTATCAGTTTTAAATACTGT
TCATTGTGTGCAGATATAAGGGGAATAGGGCATTCTGTAGAATTATACATGTCTAGTTTGTAAAGTGTGTCCTGTGT
ACTGCAGATGTGTGTTCTCTGGGCTTTATGTATCTGTACAGTAGCTTTCACATTAAAAAAATTGTGGACAAACTTGT
CCGGGGGGTTTGAGGGGAGAATGGTGGTTTATATCAATAACGATGCTGTACTATAGTCCATGTAACAAAAGATCTGG
AAGTCACCCTCCTCTGGCCCACGGAAAATTTTGGTAATCTTCTAGGTTCTAAAATGAAGATGTATGGGTACTCTGGC
AGACTGCATGTTGTATAATTTGAAAAATACTAAAAGTGGAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 69

<211> Length : 6,038

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 69

>M62096_PEA_1_T9

GCTGATTGTCCCCATGAAGGCCAGCCTTGAAGCTTGGTCAGTCTCCCTAACTGTATGATTGATCCCCACTTATTGCA
CTACATCACTGAGTTCCCGTATGCCAAGTTATGGCCACTTACATCCACGTCAGCAAAACTGGTGCCGAGGGAGCTGT
TCTTGACGAAGCTAAAAATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAA
AAACACATGTGCCATACCGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACC
ATCGTCATTTGCTGTTCTCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGGGTATG
AAATCAAGGCTTAGGTGCAAAGCCATTGGATACCATACCTGAGACCACACAGCCAAGCTAAGACCATCAAGAATACA
GTCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTT
GAAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATGGAGAAGCTGTGCCTGAGGATGAACAGA
TCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATCATAGACAATATTGCTCCTGTTGTTGCT
GGCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCTCTACAGACAACTGGATGACAAGGATGA
TGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGATGAGCTTTTAGCTTCCA
CAAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACACGTCTCCAGATTGAAAATGAGGCAGCCAAGGATGAGGTG
AAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAATTATGACCAGAAATCACAGGAAGTGGAGGATAAGACCCG
GGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAACGACTACATTGACAACCACACAGAGAGAGCTGAGCCAGC
TACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTGAGATCCTGAATTTGCTGTTGAAAGATCTGGGGGAGATA
GGTGGAATTATTGGCACCAATGATGTGAAAACTTTGGCAGATGTGAATGGAGTCATTGAGGAGGAGTTTACCATGGC
CCGCCTGTACATCAGCAAGATGAAGTCAGAGGTCAAGTCCCTGGTGAACCGCAGCAAACAGCTCGAGAGCGCCCAGA
TGGACTCCAACAGGAAGATGAATGCCAGCGAGCGGGAGCTGGCAGCCTGCCAGCTGCTCATCTCCCAGCACGAAGCC
AAGATCAAGTCTCTGACAGACTACATGCAGAACATGGAACAGAAGAGGAGGCAGCTAGAAGAGTCCCAGGACTCGCT
CAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAAAAATGCACGAAGTCAGCTTCCAGGATAAGGAGAAGGAACATC
TGACGCGGTTGCAGGATGCTGAAGAAATGAAGAAGGCGCTGGAGCAGCAGATGGAGAGCCACCGGGAAGCTCACCAG

AAGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAGCAGAAAATCATTGATGAGATTCGGGATTTGAATCAGAA
ACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAAGAGAGAGAAATGAAGC
TGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCTGGAGGAGACAGTGTCT
AGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAGTTAAAAAAAGTGTGGA
GTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAAGCAGAAAATTTCCTTCTTGGAGAATAACCTGGAGCAGC
TCACCAAAGTTCACAAGCAGCTGGTCCGGGACAACGCAGACCTGCGCTGTGAACTGCCCAAGCTGGAGAAGCGGCTG
CGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAGCGCGCTGAAGGAGGCCAAGGAGAACGCCATGCGGGACCGTAA
GCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGGCCGTGCGGGCCAAGAACATGGCCAGAAGGGCCCATTCAGCCC
AGATCGCCAAGCCCATCCGCCCCGGACACTACCCGGCCTCATCTCCAACGGCCGTCCATGCCATTCGAGGGGGAGGA
GGCAGCTCTTCAAATTCCACTCACTACCAGAAATAAATACAAAATATGACTCCACGTAGCATGTCAAGGACTACATT
AATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCCATTTTTTTTTTATACTTGCTTACTCCAGCCATCTGCAG
TACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCTGTGTGTACAGATGTGTGCTCGGACTTTTCTCTTTTTGA
GAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTACTGAGAACCATTACCACCGACTCGGCCTCCGGGGTGTT
GGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCAGTGCACTGTCCCATCTGTACGCCCTAATGTGCCATTCC
CTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATCATGGTCTGCCTCAACTGTCTGGTTTCCTGTAAAATAAA
CACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCTATAGGGTTCAAAGTTTTCCTTCTGAACACTTTTCCGAA
ACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCACATCTTTGGATCTGTAAAATATACCTTTTAGTATGGCA
CCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAAACAATCTGACAGTAGCAGTGTAGAATTTGTTCATTCAA
ATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCTCCGAGCTGCTTGCTTTTGAACCTGCAGCAACTAGTCTT
AGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTGAAAATGCTGCAAAGTTGGATAAGTGGAAATGTGGCTGC
CCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCTTGCATTGGGAATGGCTGCTTTCTCTAACCATTTTCAGC
TTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAGCAGTTGAAAAAGGAAGCCTTCGGGAGAAAGTGCTTGTC
AAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCCATGAATAAGTTATTATTTTAACCCATAATTGGCGACTG
TTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCATCTTCTTAGGCACAATTAGAAAAAATCCACATAGATGGA
TATTTTACATTTAGTTATTGCTTTATCCAAATACATGAATCTAAAGCTGAATCAACCCTTACTTCCAGTTGTGCTTA
TTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGAAACTGACTGTGCTGCTATTTGTTTTGACAAATTTGGGG
GTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTCCTATCCTATCATGTTGTGTGCCCAAGATGAGTGAGCTG
GCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATATATGCCAGGGGAGACTTTGGGCTTTTCTCATGACTGTGT
GGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTAAAGTGCTA
AGAACTGTGCATTGACATCCAAACATTTCTTGTACAAAATTTCCCTAGCAAAGCAAACCTGCTTTGACTTAATTTAT
TTGTTAAATGTTGCACTTTGTTTATGTATGTTTTGTTTTTGGTGGGGAATAAGGAGAGAGAGGACGACAAATTCTAT
TGAAGTATTTATTTTGTGAAGATGGCAATTTTGCATTTGTTTAAATTTTTTTCATTCTTTAATTTTGTTATCAGTGC
CAGCCCAATATACCTGCTCTACCATTATTTGCGGTCTGATAAAAGGGTCCTTGTGGGGCAGGTTTTGCAAAGCTTAT
CAGGTAATAACATATGCCACATAACCTTGTTGATATGTTTGCTTCTGATTTGGGAAGCTAAACATTGGTGTTTGAGA
GGATTGCCAATTATTAATTGTCATTACCACTACTCTCCATTACTTTTTGTTTGGAAATTGAACAAAGGTCAGTAATG
GTTTTTGGCTCTTGTTAATATCCATCATAAAATAGATTGTTTTAGATTCTTTCCAGGGTGATTTTTCCCTGGGTACC
CCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTCATGTTTCAAAGGGAAACATTCGCTTGTAGTTCCATTTTACTT
GATCTCTACAAGGGACTGACAACATTTGCTTTACTTTTATTCACAGAGAAAGTTGGCTTTGATGTCTCTTAAAGATA

ATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACCTAGCTTCAATCTTTATAGGACTTCTAATCTAATTTTCCTATA
GTGTGACTAAAAGGGAGGCAAATTATTGGAACGGATTATTCAAATGGATCCTTAAATATTGCTATGTATAATAAGCC
AGTTATTATATCAGGACCATGTTCTCTGTAGGCCACTTTCTAAAAAAGCCACATATGTGCAATTTTCAGGTTTTTAG
ACTATTGCTCCCTGTACTTTAAATGTAAAAACCACACTTCTGAACAACTAAGCTCATGAATATGATTTTGGTTATAT
GCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCTCCTCTATGAATAATTTTATATTTCATGCTACTTCTTGAAAGT
TTACTCTTTGATGCTCTAAGAGAACAGCCAGATGGTTTATATGAATAATCTTTATCTGCAGGATGGTGGATTGGTAA
ATTAGGAGAATGTTGTTTGAGATATCAAGATTTATGTCTGGGAACTAAAATATATAATGCCAAATGTGTTTTTGTCA
ATTACTAGAGAATTCTGTGCAAACATATCATCTCTTCAAATGCTGCACACTTTGCTTTTGTTAAACAGCAGGTAGTA
GACAGAACAATAACAGTTTCGCGTTAAGACTTTTAAAGGAAATAGAATCGTGATTAAGAAATCAGAATTTATAGATA
TATTGGGATAAATGAAGAAATAAAAATGTTTGTCTAGAATGTAGCATCTAGTGACTTTTTAAAGCCCTAACGTTTAC
ATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAGCAAATAAAAGTTCTTAACAATCCCCTCTTTCGAAGTGCATTT
TTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGCTATATGACATACTGGGAAAGGCAGAGGGTGGAGGGAAGATTT
CACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCTAAAGAAACCATCATAATCTAAAATTGCTTCATTTAACACTAA
CAATTTAGACTTTTTAAACCAAGCATTGAATAATGGCTGGATAACTGCCGAAGTAAGCGCCGCTCCATGAAGTCTGC
TTACTTATTTAAAAATTGTGTATCAGTTTTAAATACTGTTCATTGTGTGCAGATATAAGGGGAATAGGGCATTCTGT
AGAATTATACATGTCTAGTTTGTAAAGTGTGTCCTGTGTACTGCAGATGTGTGTTCTCTGGGCTTTATGTATCTGTA
CAGTAGCTTTCACATTAAAAAAATTGTGGACAAACTTGTCCGGGGGGGTTTGAGGGGAGAATGGTGGTTTATATCAAT
AACGATGCTGTACTATAGTCCATGTAACAAAAGATCTGGAAGTCACCCTCCTCTGGCCCACGGAAAATTTTGGTAAT
CTTCTAGGTTCTAAAATGAAGATGTATGGGTACTCTGGCAGACTGCATGTTGTATAATTTGAAAAATACTAAAAGTG
GAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 70

<211> Length : 5,044

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 70

>M62096_PEA_1_T11

AAGCTTCCCTGGCTTGTGCCTTTCAAATGGACTCTGGGTTTTCCTTCTGGTAGTGCATAGCTGTTCCTTTACAGGCG
CTTAGGCGTGGCTCTAGGAAAGGTTTTATGAGTCCTGGGCTGATGTAAATGTTGACCAAACACCCTCAACCAGATGG
CGAGTTTCTGTTTGCAGCAGAGCCCAGGCTGTCTTTTCTTCATAATTCTCTCTGTGCCCACTCCTCGAGGGCAGGAA
CTGTCCCTGTATCAGTGAGGCATTCGGACTTGGGAGATGTTTTTAGAACATCAGACCAGAAATGAGGGAAGGTGGAA
ATGGCCAAATCAGGTTCCCCAAGTGACTGCATGCCATCCGAGGGGCCGAGGAAGCAGAGTTCTTCTGACATGGGCTC
TCTGTTTTAAAATATCAGCCCTTCTCCCATCTCATTATTTTTCCCCTGAAGCTCTTGCACAAGCAAACTATAAATAC
ATCCTCAAAGCCTTATGTTTCATGACTCTTAGATGCACCCCAGAAATTAGTTTTCACTTGGGCACGGAGGAAGCTGT
GAGGCTGTTCTGTGATCCCTCCAAATCCTGCAGAATTACTGCCTTTATTGTACAGAGCTAATAGGGTTGGAACAGAA

CCACGGTTTTAGCCTGATGACTCAGAATTTCAGACTGATGTGGAATATATTGCTTTTTCCTCTCAATTTCAGTTTGA
ATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAAGAGAGAGAA
ATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCTGGAGGAGAC
AGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAGTTAAAAAAA
GTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAAGCAGAAAATTTCCTTCTTGGAGAATAACCTG
GAGCAGCTCACCAAAGTTCACAAGCAGCTGGTCCGGGACAACGCAGACCTGCGCTGTGAACTGCCCAAGCTGGAGAA
GCGGCTGCGTGCCACGGCGGAGCGCGTCAAGGCTCTGGAGAGCGCGCTGAAGGAGGCCAAGGAGAACGCCATGCGGG
ACCGTAAGCGCTACCAGCAGGAGGTGGATCGTATCAAGGAGGCCGTGCGGGCCAAGAACATGGCCAGAAGGGCCCAT
TCAGCCCAGATCGCCAAGCCCATCCGCCCCGGACACTACCCGGCCTCATCTCCAACGGCCGTCCATGCCATTCGAGG
GGGAGGAGGCAGCTCTTCAAATTCCACTCACTACCAGAAATAAATACAAAATATGACTCCACGTAGCATGTCAAGGA
CTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCCATTTTTTTTTTATACTTGCTTACTCCAGCCA
TCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCTGTGTGTACAGATGTGTGCTCGGACTTTTCTC
TTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTACTGAGAACCATTACCACCGACTCGGCCTCCG
GGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCAGTGCACTGTCCCATCTGTACGCCCTAATGTG
CCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATCATGGTCTGCCTCAACTGTCTGGTTTCCTGTA
AAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCTATAGGGTTCAAAGTTTTCCTTCTGAACACTT
TTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCACATCTTTGGATCTGTAAAATATACCTTTTAG
TATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAAACAATCTGACAGTAGCAGTGTAGAATTTGTT
CATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCTCCGAGCTGCTTGCTTTTGAACCTGCAGCAAC
TAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTGAAAATGCTGCAAAGTTGGATAAGTGGAAATG
TGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCTTGCATTGGGAATGGCTGCTTTCTCTAACCAT
TTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAGCAGTTGAAAAAGGAAGCCTTCGGGAGAAAGT
GCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCCATGAATAAGTTATTATTTTAACCCATAATTG
GCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCATCTTCTTAGGCACAATTAGAAAAAATCCACAT
AGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACATGAATCTAAAGCTGAATCAACCCTTACTTCCAGTT
GTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGAAACTGACTGTGCTGCTATTTGTTTTGACAAA
TTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTCCTATCCTATCATGTTGTGTGCCCAAGATGAG
TGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATATATGCCAGGGGAGACTTTGGGCTTTTCTCATG
ACTGTGTGGGTCGAAGGTAGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTAA
AGTGCTAAGAACTGTGCATTGACATCCAAACATTTCTTGTACAAAATTTCCCTAGCAAAGCAAACCTGCTTTGACTT
AATTTATTTGTTAAATGTTGCACTTTGTTTATGTATGTTTTGTTTTTGGTGGGGAATAAGGAGAGAGAGGACGACAA
ATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATTTTGCATTTGTTTAAATTTTTTTCATTCTTTAATTTTGTTA
TCAGTGCCAGCCCAATATACCTGCTCTACCATTATTTGCGGTCTGATAAAAGGGTCCTTGTGGGGCAGGTTTTGCAA
AGCTTATCAGGTAATAACATATGCCACATAACCTTGTTGATATGTTTGCTTCTGATTTGGGAAGCTAAACATTGGTG
TTTGAGAGGATTGCCAATTATTAATTGTCATTACCACTACTCTCCATTACTTTTTGTTTGGAAATTGAACAAAGGTC
AGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAAAATAGATTGTTTTAGATTCTTTCCAGGGTGATTTTTCCCT
GGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCACTTTCATGTTTCAAAGGGAAACATTCGCTTGTAGTTCCAT
TTTACTTGATCTCTACAAGGGACTGACAACATTTGCTTTACTTTTATTCACAGAGAAAGTTGGCTTTGATGTCTCTT

AAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGTTCACCTAGCTTCAATCTTTATAGGACTTCTAATCTAATTT
TCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGAACGGATTATTCAAATGGATCCTTAAATATTGCTATGTATA
ATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTAGGCCACTTTCTAAAAAAGCCACATATGTGCAATTTTCAGG
TTTTTAGACTATTGCTCCCTGTACTTTAAATGTAAAAACCACACTTCTGAACAACTAAGCTCATGAATATGATTTTG
GTTATATGCAGCTTTTGACTAGCATGTATTGTGTCTTTTTCTCCTCTATGAATAATTTTATATTTCATGCTACTTCT
TGAAAGTTTACTCTTTGATGCTCTAAGAGAACAGCCAGATGGTTTATATGAATAATCTTTATCTGCAGGATGGTGGA
TTGGTAAATTAGGAGAATGTTGTTTGAGATATCAAGATTTATGTCTGGGAACTAAAATATATAATGCCAAATGTGTT
TTTGTCAATTACTAGAGAATTCTGTGCAAACATATCATCTCTTCAAATGCTGCACACTTTGCTTTTGTTAAACAGCA
GGTAGTAGACAGAACAATAACAGTTTCGCGTTAAGACTTTTAAAGGAAATAGAATCGTGATTAAGAAATCAGAATTT
ATAGATATATTGGGATAAATGAAGAAATAAAAATGTTTGTCTAGAATGTAGCATCTAGTGACTTTTTAAAGCCCTAA
CGTTTACATAAAGAAGCTCTAGTTCTTATAGAAATAACAAAGCAAATAAAAGTTCTTAACAATCCCCTCTTTCGAAG
TGCATTTTTTTAAAGCAGGGCAGGAGACATTTGGACTCTAGCTATATGACATACTGGGAAAGGCAGAGGGTGGAGGG
AAGATTTCACTTCATTGTCTAGCCCAGAATCTTGAGCAAGCTAAAGAAACCATCATAATCTAAAATTGCTTCATTTA
ACACTAACAATTTAGACTTTTTAAACCAAGCATTGAATAATGGCTGGATAACTGCCGAAGTAAGCGCCGCTCCATGA
AGTCTGCTTACTTATTTAAAAATTGTGTATCAGTTTTAAATACTGTTCATTGTGTGCAGATATAAGGGGAATAGGGC
ATTCTGTAGAATTATACATGTCTAGTTTGTAAAGTGTGTCCTGTGTACTGCAGATGTGTGTTCTCTGGGCTTTATGT
ATCTGTACAGTAGCTTTCACATTAAAAAAATTGTGGACAAACTTGTCCGGGGGGGTTTGAGGGGAGAATGGTGGTTTA
TATCAATAACGATGCTGTACTATAGTCCATGTAACAAAAGATCTGGAAGTCACCCTCCTCTGGCCCACGGAAAATTT
TGGTAATCTTCTAGGTTCTAAAATGAAGATGTATGGGTACTCTGGCAGACTGCATGTTGTATAATTTGAAAAATACT
AAAAGTGGAAAATAAAATTGAATTAAACTTTGGCTGGTC

<210> SEQ ID NO 71

<211> Length : 2,945

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 71

>M62096_PEA_1_T13

CGGCAGAGCGCGCTGGTGCTGATGCAGGATGGCTGAGCGCGCAGGAGCCCGGGAGGTCTGAGCCGGGCGAGGCTCGC
TCCCTGCGCATCGCCTCCTCCGCCCGCCGCGTGGTCGCGGGCAGGTGGGCCGGGGGGCGCTGGGCAGGGGCGGGGCA
GGGCCAGGGCAGGCCGGTCTGCAGCCGGAGGGGCCGGAGCGGAGAAGCTGCCCACCTTCCCGGGCTCGGAGCGGCCG
GGGCTGCTCAGCCGGCCGGGCTCGCGATGACCTGCTGAGAAGCGTCGTCGGAGGCTGCAGGAGGCGGCCTAGCTGTG
GGCGGTGCAGCTCGCGGCCTCCTCCCTCGTCGTTCCCGGCCCCGGCCCCCCACCCATCCCCGTGCCCCCTCCCTACC
GCCGGCCGAGATGGCGGATCCAGCCGAATGCAGCATCAAAGTGATGTGCCGGTTCCGGCCCCTCAACGAAGCGGAGA
TCCTCCGCGGGGACAAATTCATCCCCAAATTTAAAGGCGATGAGACCGTGGTGATCGGGCAAGGGAAGCCATATGTC
TTCGACAGAGTGCTACCTCCCAACACGACCCAAGAGCAGGTTTACAATGCATGTGCGAAGCAAATTGTCAAAGATGT

CCTTGAAGGTTATAACGGGACGATTTTTGCGTATGGGCAGACTTCATCAGGAAAAACCCACACCATGGAGGGGAAGC
TGCATGACCCCCAGCTCATGGGGATCATCCCACGAATTGCCCATGATATCTTTGACCATATCTACTCCATGGATGAG
AACCTGGAGTTTCACATAAAGGTTTCCTATTTTGAGATCTACTTGGACAAAATAAGGGACTTACTTGATGTATCCAA
GACCAACTTGGCTGTTCATGAAGATAAAAACAGAGTCCCGTATGTAAAGGGGTGCACTGAGCGGTTTGTGTCGAGCC
CTGAGGAAGTCATGGATGTAATAGATGAAGGCAAAGCAAACCGACACGTGGCTGTGACAAACATGAATGAACACAGC
TCTAGAAGTCACAGTATCTTCCTGATAAATATTAAACAAGAGAATGTAGAGACTGAAAAAAAACTCAGTGGGAAACT
TTATTTGGTTGATTTGGCTGGGAGCGAAAAGGTCAGCAAAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAA
ATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAAAAACACATGTGCCATAC
CGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACCATCGTCATTTGCTGTTC
TCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGAGCTAAGACCATCAAGAATACAG
TCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTTG
AAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATGGAGAAGCTGTGCCTGAGGATGAACAGAT
CAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATCATAGACAATATTGCTCCTGTTGTTGCTG
GCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCTCTACAGACAACTGGATGACAAGGATGAT
GAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGATGAGCTTTTAGCTTCCAC
AAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACACGTCTCCAGATTGAAAATGAGGCAGCCAAGGATGAGGTGA
AAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAATTATGACCAGAAATCACAGGAAGTGGAGGATAAGACCCGG
GCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAACGACTACATTGACAACCACACAGAGAGAGCTGAGCCAGCT
ACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTGAGATCCTGAATTTGCTGTTGAAAGATCTGGGGGAGATAG
GTGGAATTATTGGCACCAATGATGTGAAAACTTTGGCAGATGTGAATGGAGTCATTGAGGAGGAGTTTACCATGGCC
CGCCTGTACATCAGCAAGATGAAGTCAGAGGTCAAGTCCCTGGTGAACCGCAGCAAACAGCTCGAGAGCGCCCAGAT
GGACTCCAACAGGAAGATGAATGCCAGCGAGCGGGAGCTGGCAGCCTGCCAGCTGCTCATCTCCCAGCACGAAGCCA
AGATCAAGTCTCTGACAGACTACATGCAGAACATGGAACAGAAGAGGAGGCAGCTAGAAGAGTCCCAGGACTCGCTC
AGCGAAGAGCTGGCAAAGCTCCGAGCCCAGGAAAAAATGCACGAAGTCAGCTTCCAGGATAAGGAGAAGGAACATCT
GACGCGGTTGCAGGATGCTGAAGAAATGAAGAAGGCGCTGGAGCAGCAGATGGAGAGCCACCGGGAAGCTCACCAGA
AGCAGCTGTCCAGACTCCGAGACGAAATTGAGGAGAAGCAGAAAATCATTGATGAGATTCGGGAGTGAGTCGGCCCA
GGGCACCAGGGGTGTGGGGGTGGTCATGCCTCGGTCCTCTTGGGGAAGCCTGGAAGGATGTGGCTCTTAGTCGAGGG
CCCTGCTCACCTTGCCCTGTGGGCACTGCCCTGGTGACACAGCAGGCTGGGCGGGCTGCTTCCAAGGTCTGTTCTCC
GATCTGGAGCTGAGCCTCCTGGAGCCCTGGCATGGCAGGTGGCAGGCGGGCCCAGCTGCCTCTCCTAGTCCCCGAGG
GCCAGGGTCACATAGGTGATTCCGCTGGATGGACGCATGCCCCAGTAGATGGGGGGGAGCATCTGTAATGAAAGCAC
CAAAAAAAAAAAAAAAAAA

<210> SEQ ID NO 72

<211> Length : 2,261

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 72

>M62096_PEA_1_T14

CGGCAGAGCGCGCTGGTGCTGATGCAGGATGGCTGAGCGCGCAGGAGCCCGGGAGGTCTGAGCCGGGCGAGGCTCGC
TCCCTGCGCATCGCCTCCTCCGCCCGCCGCGTGGTCGCGGGCAGGTGGGCCGGGGGGCGCTGGGCAGGGGCGGGGCA
GGGCCAGGGCAGGCCGGTCTGCAGCCGGAGGGGCCGGAGCGGAGAAGCTGCCCACCTTCCCGGGCTCGGAGCGGCCG
GGGCTGCTCAGCCGGCCGGGCTCGCGATGACCTGCTGAGAAGCGTCGTCGGAGGCTGCAGGAGGCGGCCTAGCTGTG
GGCGGTGCAGCTCGCGGCCTCCTCCCTCGTCGTTCCCGGCCCCGGCCCCCCACCCATCCCCGTGCCCCCTCCCTACC
GCCGGCCGAGATGGCGGATCCAGCCGAATGCAGCATCAAAGTGATGTGCCGGTTCCGGCCCCTCAACGAAGCGGAGA
TCCTCCGCGGGGACAAATTCATCCCCAAATTTAAAGGCGATGAGACCGTGGTGATCGGGCAAGGGAAGCCATATGTC
TTCGACAGAGTGCTACCTCCCAACACGACCCAAGAGCAGGTTTACAATGCATGTGCGAAGCAAATTGTCAAAGATGT
CCTTGAAGGTTATAACGGGACGATTTTTGCGTATGGGCAGACTTCATCAGGAAAAACCCACACCATGGAGGGGAAGC
TGCATGACCCCCAGCTCATGGGGATCATCCCACGAATTGCCCATGATATCTTTGACCATATCTACTCCATGGATGAG
AACCTGGAGTTTCACATAAAGGTTTCCTATTTTGAGATCTACTTGGACAAAATAAGGGACTTACTTGATGTATCCAA
GACCAACTTGGCTGTTCATGAAGATAAAAACAGAGTCCCGTATGTAAAGGGGTGCACTGAGCGGTTTGTGTCGAGCC
CTGAGGAAGTCATGGATGTAATAGATGAAGGCAAAGCAAACCGACACGTGGCTGTGACAAACATGAATGAACACAGC
TCTAGAAGTCACAGTATCTTCCTGATAAATATTAAACAAGAGAATGTAGAGACTGAAAAAAAACTCAGTGGGAAACT
TTATTTGGTTGATTTGGCTGGGAGCGAAAAGGTCAGCAAAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAA
ATATCAATAAGTCTTTGTCTGCTCTTGGAAATGTGATCTCTGCTTTGGCAGAAGGGACAAAAACACATGTGCCATAC
CGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCACCATCGTCATTTGCTGTTC
TCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAGAGCTAAGACCATCAAGAATACAG
TCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAGAGAAAGAGAAAAACAAGACTTTG
AAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATGGAGAAGCTGTGCCTGAGGATGAACAGAT
CAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATCATAGACAATATTGCTCCTGTTGTTGCTG
GCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCTCTACAGACAACTGGATGACAAGGATGAT
GAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGATGAGGTAAAGAATGCAAT
ATATTTTTTTTTCCACAAAGTTCTTCTATTACTCTTTGTTGTTGATGTTTGTTCCAGGAATTTAATTGGCATAGAAG
CTTTTCATAATTACAGAATCATGTGGAAATTTCTTGGTAGATGTCCCTTCACTGCCTCTTACAAGCTGATTATCACT
GAATTTAGAAAATAAATGTCTGACTTTCAAAAACCCCTGATGTTTTGAGATTGAGTAGCCAGTGGCTACAGTTCGTT
CTGGAAGGGCAGAGACCTTTGGTTGGGTGATCAAGCAAGGATGATCCTTTTTTATTTTTATTTTTTTGAGACAGGGT
CTCTCTGTTGTCCAGGCTGGAATGCAGTGGTGCAATCATGGCTCACTGCAACCTCCAGAGCTCAAATGATCTTCCCG
CCTAAGACTCTCAAGTAGCTAAGACTACAAGAATGTGCCACCATACCTAGCTAATTTTTTAATATTTTGAGACAGAG
TTTCTCTATGTTGGTCAGGGTGATCTTG

<210> SEQ ID NO 73

<211> Length : 1,059

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 73

>M62096_PEA_1_T15

AAGCTTCCCTGGCTTGTGCCTTTCAAATGGACTCTGGGTTTTCCTTCTGGTAGTGCATAGCTGTTCCTTTACAGGCG
CTTAGGCGTGGCTCTAGGAAAGGTTTTATGAGTCCTGGGCTGATGTAAATGTTGACCAAACACCCTCAACCAGATGG
CGAGTTTCTGTTTGCAGCAGAGCCCAGGCTGTCTTTTCTTCATAATTCTCTCTGTGCCCACTCCTCGAGGGCAGGAA
CTGTCCCTGTATCAGTGAGGCATTCGGACTTGGGAGATGTTTTTAGAACATCAGACCAGAAATGAGGGAAGGTGGAA
ATGGCCAAATCAGGTTCCCCAAGTGACTGCATGCCATCCGAGGGGCCGAGGAAGCAGAGTTCTTCTGACATGGGCTC
TCTGTTTTAAAATATCAGCCCTTCTCCCATCTCATTATTTTTCCCCTGAAGCTCTTGCACAAGCAAACTATAAATAC
ATCCTCAAAGCCTTATGTTTCATGACTCTTAGATGCACCCCAGAAATTAGTTTTCACTTGGGCACGGAGGAAGCTGT
GAGGCTGTTCTGTGATCCCTCCAAATCCTGCAGAATTACTGCCTTTATTGTACAGAGCTAATAGGGTTGGAACAGAA
CCACGGTTTTAGCCTGATGACTCAGAATTTCAGACTGATGTGGAATATATTGCTTTTTCCTCTCAATTTCAGTTTGA
ATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAAGAGAGAGAA
ATGAAGCTGGAAAAGCTCTTATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCTGGAGGAGAC
AGTGTCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAGTTAAAAAAG
TGAGTTCTCTTTGTCTGAATGGGACTGAGAAGAAAATCAAAGATGGCAGGGAAGAATCATTTTCAGTTGAAATATCA
CTTGCTTAAGTCGGGGCTGGTTATGCTTAAAAATTAATTACTGCACACCAGAGAATGT

<210> SEQ ID NO 74

<211> Length : 2,715

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 74

>M78076_PEA_1_T2

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA

CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCC
AGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGG
TGAGTGAGCCCACATATAGATGACCCCAGACATTAGGGAACAGGCCCCAGCCTAATTTGTAATCCCCTAGAGTCTGA
GGGTGTCTTCACCACCACAGTGACTGGGAGAGGATGAGGAGGAACGTCTAAGGTTGCAGGGGCCTCTGTAGGATCCC
CAATCCTCCTTCTTAGTCCCTGGAAGGATGTTTCTCCACCTTTCTTTGCTGATACCCTCCTCTCTTCACTGTTCCAC
TCCCTTGCTTCCTCTGGCTGCCAGCAGACACCCCCATGACCCTTCCAAAAGGGTCCACAGAACAAGATGCTGCATCC
CCTGAGAAAGAGAAGATGAACCCGCTGGAACAGTATGAGCGAAAGGTGAATGCGTCTGTTCCAAGGGGTTTCCCTTT
CCACTCATCGGAGATTCAGAGGGATGAGCTGGCACCAGCTGGGACAGGGGTGTCCCGTGAGGCTGTGTCGGGTCTGC
TGATCATGGGAGCGGGCGGAGGCTCCCTCATCGTCCTCTCCATGCTGCTCCTGCGCAGGAAGAAGCCCTACGGGGCT
ATCAGCCATGGCGTGGTGGAGGTGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTCCGCGAACTGCAGCGGCACGG
CTATGAGAACCCCACTTACCGCTTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCCTTCAGCCGAGCCCAGA
CCTCCCCTCTTCCTGGAGCCCCAGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTTCACA
CCCTTTTGTGAGACGGCTGGAAATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAGATAG
TCCCAGCAGCCTCCCCACGTGGCACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGTGACC
GCCACCTTGGTCCTAGTGTCTATTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTCCTCT
TCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 75

<211> Length : 2,931

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 75

>M78076_PEA_1_T3

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCC
AGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGA
CACCCCCATGACCCTTCCAAAAGGTGAGTGTCTCACAGTTAACCCCAGCCTCCAAATCCCACTGAATCCCTGAACCC
AGAAGGAAACAGGGTCCATCCATTGGGAACCTCAGACCCCCTGGGGGTAGAGTTTGATGTACTTTCCAGCCCCCTCCT
CTGGACCCTAAAGAATGAGATAGGGCCAGGCGCTGGTGACTCACACCCGTAATCCTAGCACTTTCAGAGGCTGAGGC
AGGAGGATCCCTTGAGGCCACGAGTTCTAGACCAGCCTGGGCAACATAATGAGACCCTGTACCTACAAATAATTTAA
AAATTACCTGGGTGTGGTGGGGCATGTCTGTAGTCCCAGCTGCTCAGGAGGCTGACGTAGAAGGATCACTGGAGCCC
AGGAAGTTGAGGCTGCAGTGAGCTGAGATCATGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCTGTCTAAA
GAAAAAAAAAAAGAATGAGATCAGACTTGGGGGTAGGGTCCACAGAACAAGATGCTGCATCCCCTGAGAAAGAGAAG
ATGAACCCGCTGGAACAGTATGAGCGAAAGGTGAATGCGTCTGTTCCAAGGGGTTTCCCTTTCCACTCATCGGAGAT
TCAGAGGGATGAGCTGGCACCAGCTGGGACAGGGGTGTCCCGTGAGGCTGTGTCGGGTCTGCTGATCATGGGAGCGG
GCGGAGGCTCCCTCATCGTCCTCTCCATGCTGCTCCTGCGCAGGAAGAAGCCCTACGGGGCTATCAGCCATGGCGTG
GTGGAGGTGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTCCGCGAACTGCAGCGGCACGGCTATGAGAACCCCAC
TTACCGCTTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCCTTCAGCCGAGCCCAGACCTCCCCTCTTCCTG
GAGCCCCAGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTTCACACCCTTTTGTGAGACG
GCTGGAAATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAGATAGTCCCAGCAGCCTCCC
CACGTGGCACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGTGACCGCCACCTTGGTCCTA

GTGTCTATTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTCCTCTTCCCTACCAGGCCAG
TCTGA

<210> SEQ ID NO 76

<211> Length : 3,190

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 76

>M78076_PEA_1_T5

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCC
AGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGG
TGAGTGAGCCCACATATAGATGACCCCAGACATTAGGGAACAGGCCCCAGCCTAATTTGTAATCCCCTAGAGTCTGA
GGGTGTCTTCACCACCACAGTGACTGGGAGAGGATGAGGAGGAACGTCTAAGGTTGCAGGGGCCTCTGTAGGATCCC
CAATCCTCCTTCTTAGTCCCTGGAAGGATGTTTCTCCACCTTTCTTTGCTGATACCCTCCTCTCTTCACTGTTCCAC
TCCCTTGCTTCCTCTGGCTGCCAGCAGACACCCCCATGACCCTTCCAAAAGGTGAGTGTCTCACAGTTAACCCCAGC

CTCCAAATCCCACTGAATCCCTGAACCCAGAAGGAAACAGGGTCCATCCATTGGGAACCTCAGACCCCCTGGGGTAG
AGTTTGATGTACTTTCCAGCCCCCTCCTCTGGACCCTAAAGAATGAGATAGGGCCAGGCGCTGGTGACTCACACCCG
TAATCCTAGCACTTTCAGAGGCTGAGGCAGGAGGATCCCTTGAGGCCACGAGTTCTAGACCAGCCTGGGCAACATAA
TGAGACCCTGTACCTACAAATAATTTAAAAATTACCTGGGTGTGGTGGGGCATGTCTGTAGTCCCAGCTGCTCAGGA
GGCTGACGTAGAAGGATCACTGGAGCCCAGGAAGTTGAGGCTGCAGTGAGCTGAGATCATGCCACTGCACTCCAGCC
TGGGTGACAGAGTGAGACTCTGTCTAAAGAAAAAAAAAAAGAATGAGATCAGACTTGGGGGTAGGGTCCACAGAACA
AGATGCTGCATCCCCTGAGAAAGAGAAGATGAACCCGCTGGAACAGTATGAGCGAAAGGTGAATGCGTCTGTTCCAA
GGGGTTTCCCTTTCCACTCATCGGAGATTCAGAGGGATGAGCTGGCACCAGCTGGGACAGGGGTGTCCCGTGAGGCT
GTGTCGGGTCTGCTGATCATGGGAGCGGGCGGAGGCTCCCTCATCGTCCTCTCCATGCTGCTCCTGCGCAGGAAGAA
GCCCTACGGGGCTATCAGCCATGGCGTGGTGGAGGTGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTCCGCGAAC
TGCAGCGGCACGGCTATGAGAACCCCACTTACCGCTTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCCTTC
AGCCGAGCCCAGACCTCCCCTCTTCCTGGAGCCCCAGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAG
TGATCATTTCACACCCTTTTGTGAGACGGCTGGAAATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAG
AATTCCCAGATAGTCCCAGCAGCCTCCCCACGTGGCACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCT
CTTTAAGGTGACCGCCACCTTGGTCCTAGTGTCTATTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCC
AATAAAGTCCTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 77

<211> Length : 2,385

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 77

>M78076_PEA_1_T13

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG

CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCC
AGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGA
CACCCCCATGACCCTTCCAAAAGGTGAATGCGTCTGTTCCAAGGGGTTTCCCTTTCCACTCATCGGAGATTCAGAGG
GATGAGCTGGCACCAGCTGGGACAGGGGTGTCCCGTGAGGCTGTGTCGGGTCTGCTGATCATGGGAGCGGGCGGAGG
CTCCCTCATCGTCCTCTCCATGCTGCTCCTGCGCAGGAAGAAGCCCTACGGGGCTATCAGCCATGGCGTGGTGGAGG
TGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTCCGCGAACTGCAGCGGCACGGCTATGAGAACCCCACTTACCGC
TTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCCTTCAGCCGAGCCCAGACCTCCCCTCTTCCTGGAGCCCC
AGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTTCACACCCTTTTGTGAGACGGCTGGAA
ATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAGATAGTCCCAGCAGCCTCCCCACGTGG
CACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGTGACCGCCACCTTGGTCCTAGTGTCTA
TTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTCCTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 78

<211> Length :

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 78

>M78076_PEA_1_T15

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG

GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCC
AGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGA
CACCCCCATGACCCTTCCAAAAGGGTCCACAGAACAAGATGCTGCATCCCCTGAGAAAGAGAAGATGAACCCGCTGG
AACAGTATGAGCGAAAGGTGAATGCGTCTGTTCCAAGGGGTTTCCCTTTCCACTCATCGGAGATTCAGAGGGATGAG
CTGGTAAGAGGAGGAACAGCCGGGTACCTAGGGGAAGAGACCAGAGGTCAGCGGCCAGGCTGTGATTCCCAAAGCCA
CACAGGACCCTCAAAGAAGCCCTCTGCCCCATCTCCTCTCCCTGCAGGCACCAGCTGGGACAGGGGTGTCCCGTGAG
GCTGTGTCGGGTCTGCTGATCATGGGAGCGGGCGGAGGCTCCCTCATCGTCCTCTCCATGCTGCTCCTGCGCAGGAA
GAAGCCCTACGGGGCTATCAGCCATGGCGTGGTGGAGGTGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTCCGCG
AACTGCAGCGGCACGGCTATGAGAACCCCACTTACCGCTTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCC
TTCAGCCGAGCCCAGACCTCCCCTCTTCCTGGAGCCCCAGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTG
AAGTGATCATTTCACACCCTTTTGTGAGACGGCTGGAAATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCT
AAGAATTCCCAGATAGTCCCAGCAGCCTCCCCACGTGGCACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATG
GCTCTTTAAGGTGACCGCCACCTTGGTCCTAGTGTCTATTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACAT
CCCAATAAAGTCCTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 79

<211> Length : 2,297

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 79

>M78076_PEA_1_T23

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT

CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGCTACCT
GCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAAGGCAC
AGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGCTTGAC
CAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCAGGAACTCCTCCACTCTGAACACCTGGGTCCCAGTGA
ATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAGGGTGGGCTGCAGCCTCCAGATTCCAAGGATGACACCC
CCATGACCCTTCCAAAAGGGTCCACAGAACAAGATGCTGCATCCCCTGAGAAAGAGAAGATGAACCCGCTGGAACAG
TATGAGCGAAAGGTGAATGCGTCTGTTCCAAGGGGTTTCCCTTTCCACTCATCGGAGATTCAGAGGGATGAGCTGGC
ACCAGCTGGGACAGGGGTGTCCCGTGAGGCTGTGTCGGGTCTGCTGATCATGGGAGCGGGCGGAGGCTCCCTCATCG
TCCTCTCCATGCTGCTCCTGCGCAGGAAGAAGCCCTACGGGGCTATCAGCCATGGCGTGGTGGAGGTGGACCCCATG
CTGACCCTGGAGGAGCAGCAGCTCCGCGAACTGCAGCGGCACGGCTATGAGAACCCCACTTACCGCTTCCTGGAGGA
ACGACCCTGACCCGGCCCCCTTCACCCCTTCAGCCGAGCCCAGACCTCCCCTCTTCCTGGAGCCCCAGAACCCCAAC
TCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTTCACACCCTTTTGTGAGACGGCTGGAAATTCTTATTTC
CCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAGATAGTCCCAGCAGCCTCCCCACGTGGCACCTCCTCAC
CTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGTGACCGCCACCTTGGTCCTAGTGTCTATTCCCTGGAAT
TCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTCCTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 80

<211> Length : 2,457

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 80

>M78076_PEA_1_T26

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGC
TTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCGTGAGTGTCTATTACCCTGGCTCCCATTACAG
ATCTCTGAGGGCAGATCTTGACTCCTAAATGTTGGGCCCCCCCAATTTCATTTATTCCTCTATAACAAACAGCCCAG
ACCTTAGCAGTGAAAATCAACAATGATTTTTCTTTGTTCATGATTCTGCCATCCGGTCTGCGCTCAGCAGAGTGGTT
CTTTCAGTGGTCTTGCCAGTGGTCAAGCATGCAGCTGTATTTAGCTAGCAGATCATCTAGGGGCTGGGAGTCTAGCA
CAAATGGACCTTTCTCTCTCTCCAAGGAAGCGCAAGGCCTCTCTTCTCCGTGGAGCTTCTCCATGTGGTCTCATCAG
CAGGGTAGCTAGATTCCCTACATGGTGGTTTATGCTCTCTAAGACATCACAGTGGAAGTTGCTAGGTCTTAAGGCTT
GGGCCCACATTCTATTTGTTAAAGCAAGTTACAAATTCAGTCCAGATTCAAGGGAAGGAACCTATATGCATACCGGA
AAGTGTGACCTATTGCAGCCCCCACATCTATTGTGTCTTTCTCCTGGATATCTCACACATAACCCTGATTCTCCTAG
TATTTAAGAAAGCTATCATCTTGAGGCGCGGTGGCTCACGCCTATAATCCCAGCACTTTAGGAGGCCGAGGCGGGTG
GATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTTTACTAAAAATACAAAAATC
AGCCGGGCATGATGTCGCTTGCCTGTAATCCCAGCTACTTAGGAGGCTGAGGCAAGAGAATTGCTTGAACCCGGGAG
GTGGAGGTTGCAGTGAGCTGAGATCGCATCATTGCACTCCAGCTGGGCAACAAGAGTGAGACTCTGTCTC

<210> SEQ ID NO 81

<211> Length : 4,104

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 81

>M78076_PEA_1_T27

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA
CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGGTGCTCACATCCTTCCAGCTCCCAAATGCGCCGCTATTCCTCAGACGCCCGCGCC
TCAGGCTCTTCTCTTGTCCCTTAGACCCTCTTTCTGTCTCTTGGACCCCTTCCTATCCCCTGAACACCGCTTCTCTG
CCCCTTCCCAGTCTCTCAGCTCAGCTTCCTGACCCTGAAACATGGACCCTCACATGCTGTGTCTTTGACCCCTGCTT
CTTGGCCCTTGGATTCCTACTCCCCCCGCCGTCGATCCTATGTTCTGTCCCTTGGATTTTCACTGCCTTTCCCAGAA
TCGTCTTTTTTTTTTTTTTTTTTTTTGAGACAGGTTCTTGCTCTGTCGCCCAGGCAGGAGAGCAGTGTGCGATCTTGG
CTCATTGCAACTTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCTCGAGTAGCTGGGATTACAGGAGCC
TGCCACCACACTGGGCTAATTTTTTTTTTTTTTTTTGACAGAGTCTCGCTCTGTTTCCCAGGCTGGAGTGCAGTGAC
ATGATCTGGGCTCACTGCAACCTCCGCCTACTGGGTTCAAGCTATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGAC
TACAGGCGGGTGTCACCACATCTGGCTGATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATACTGGTCAGGCTG
GTCTTGAACTCGACCTCAGGTGATCCACCCTTGGCCTCCTAAAGTACTCGGATTACAGGTGTGAGCCACCACGCCCG
GCCCCAGCTAATTTTTGTATTTTTGGTAGACACGGGTTTCAGCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTC
AGGTGATCTGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCAGCCAGAAACCCCAA
TAACTTTTGCACCAATCTAATATTTTTAGCAGAGACAGGGTTTTGCCATGTTGCCCAGGCTGGTCTCGAACTCCTGA
CCTCAGGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCGGCCAGAAACC

CCAATAACTTGCACCAATCTAATATTTTTAGCAGAGACAGGGTTTTGCCATGTTGCCCAGGCTAGTCTCAAACTCCT
GACCTCAGGTGATCTGCCTACCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGTCGAGAA
TCTCCTTCTTGTTCCTTGAACCCTCTTCCTGTCCCTCAACCTCCTTTCTCCATAACTTCACTTGTTTTCCCTGGAAC
CCCTGTTCTGTGCGCTCAAATTTGAATTCCCCTTTCCTGGATGTTTTCTTCCTGTCTATGAAACTCCATTCTGTGCT
CTTGAACTCCAAATCTTGCCTTGAACCATGTCATTTCTATATGACCCTCCAATCCTCAATCTCTGTCTCTGGAATCC
CCTCAAACCCCACTTTCTGTTCCTTGGACTTTATTCTTCAATTTCCTTCTCCTATGGCCCAGTTCCTAACCCTTGTA
CCACACATCCTGTCCATTGCATGTGCCGCTTTTCCTCAGTCGCTATTGAATTCCTCCTTCATACTGCTTCAGTTTCC
TCATCTCCAGCCTGCATTGCGCAGTTCATCCTTCATGTCCACTCACCCACAGGTGCATACCCACCTTCAAGTGATTG
AGGAGAGGGTGAATCAGAGCCTGGGCCTGCTTGACCAGAACCCCCACCTGGCTCAGGAGCTGCGGCCCCAAATCCGT
GAGTGTCTATTACCCTGGCTCCCATTACAGATCTCTGAGGGCAGATCTTGACTCCTAAATGTTGGGCCCCCCCAATT
TCATTTATTCCTCTATAACAAACAGCCCAGACCTTAGCAGTGAAAATCAACAATGATTTTTCTTTGTTCATGATTCT
GCCATCCGGTCTGCGCTCAGCAGAGTGGTTCTTTCAGTGGTCTTGCCAGTGGTCAAGCATGCAGCTGTATTTAGCTA
GCAGATCATCTAGGGGCTGGGAGTCTAGCACAAATGGACCTTTCTCTCTCTCCAAGGAAGCGCAAGGCCTCTCTTCT
CCGTGGAGCTTCTCCATGTGGTCTCATCAGCAGGGTAGCTAGATTCCCTACATGGTGGTTTATGCTCTCTAAGACAT
CACAGTGGAAGTTGCTAGGTCTTAAGGCTTGGGCCCACATTCTATTTGTTAAAGCAAGTTACAAATTCAGTCCAGAT
TCAAGGGAAGGAACCTATATGCATACCGGAAAGTGTGACCTATTGCAGCCCCCACATCTATTGTGTCTTTCTCCTGG
ATATCTCACACATAACCCTGATTCTCCTAGTATTTAAGAAAGCTATCATCTTGAGGCGCGGTGGCTCACGCCTATAA
TCCCAGCACTTTAGGAGGCCGAGGCGGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGA
AACCCCGTCTTTACTAAAAATACAAAAATCAGCCGGGCATGATGTCGCTTGCCTGTAATCCCAGCTACTTAGGAGGC
TGAGGCAAGAGAATTGCTTGAACCCGGGAGGTGGAGGTTGCAGTGAGCTGAGATCGCATCATTGCACTCCAGCTGGG
CAACAAGAGTGAGACTCTGTCTC

<210> SEQ ID NO 82

<211> Length : 1,795

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 82

>M78076_PEA_1_T28

CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCGCTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCGCTGCTGCTGCCACTATTGCTGCTGCTT
CTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAGGCCCCGGGGTCGGCCCAGGT
GGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGAACCAGACCCACAGCGCTCTC
GACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAGATGTACCCGGAGCTGCAGATTGCACGTGTG
GAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTCCCGGAGCGGCAGCTGCGCCCACCCCCACCA

CCAGGTTGTGCCCTTCCGCTGCCTGCCTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCT
TGCACCAGGAGCGCATGGACCAATGTGAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAGGCCTGCAGCTCCCAG
GGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGTGGAGTATGTGTGCTG
TCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGGTGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCA
GAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCCCACAGCCAGTAGATGATTACTTCGTGGAGCCTCCG
CAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCCCATACACTTGCAGTGGTCGGCAAAGTCACTCCCAC
CCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGGTTCCTGAGGG
CCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAGGTGATGCGTGAATGGGCCATGGCAGACAACCAG
TCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAGCACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCA
GGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCACCCGCGTCATCGCCCTTATCAACGACCAGCGCCGGG
CTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGCCTCAGGCGGAGCGTGTCCTGTTGGCCCTGCGGCGC
TACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTACCAGCATGTGGCCGCCGTGGATCCCGAGAA
GGCACAGCAGATGCGCTTCCAGCCCCAGAACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTT
CACACCCTTTTGTGAGACGGCTGGAAATTCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAG
ATAGTCCCAGCAGCCTCCCCACGTGGCACCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGT
GACCGCCACCTTGGTCCTAGTGTCTATTCCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTC
CTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 83

<211> Length : 2,175

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 83

>T99080_PEA_4_T0

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGCTATTTACCATATAAGATAACTCTTAAGAACTGGAGATAGTCAGCTCCCCTGGGTTAATTTGA
AGCAGAAGAGGGCAGTTGTTATACTGCCCTGTCAGTTGGATGCGGAGTCTTACTCAAAATTCATTCTCAGCATTCTT
CTTTTATGGTATCTTCTTTGGCACTTAGCAGCGCATCAGGTAGGCATCTTCTATTTTTCTTCATTCCTTAATTTCCT
TTGTATCCCTCAAATGGTTATTTATTTGGCTGGAGTCTGTTTTGTTCATTAAGCAAACATGTCTTTGCTTTGAACAT
GTCTTTGATTTGATGGATACTTAAATTCCTCATCAAACATTTTGTTGCTATGCATAACGTTTTCTTTGGCCAACTCC

AGCAATTTCCCACATTTTGACATGCAATCATGTTAACTCCCATTTTCTTTTGTAATCCAACATCTTCTATTTAGATA
ATTACTTTAACAATCAATGACTTAATATTCTAATCATAAATTTATACAAAAATAAAATTACCTCCAAAACATTGCTA
CCTTTCCTAAACATTCAGTCTTGCCACAGTTTAATAAAAGGAAAGAACATTAAAAAGGATAAGACACTGTAATGATT
AGATGCTTTTTATAAGCCTAAAGGCATTGTGATTATTTAGACAGAAGAGAAGAAAGTGAAGTGAAAACCTGATAGTT
ATGTAGTCTCATGGTTTGCTGTTGAGAGGCTGAACACCAGCTGCTTTCCTTTTCTAGGAAGATAATAAAGTGGGCTT
TGGCTACAACATAAAGATGTTGGGTTAGACAGTTTCACTACAGTAAGAACAACGGGATGAGTTGCCCAGGAAATTGT
GAAATACTTTCTAATGATCTTTAAAGATATAATGAACACTAATTCATCTGGATTTGTTTAGGTGTGGTCCTGGTTAA
AGGCAAAGGGAAGGATCAGATAACTTCATGTTTTTTCCATTTAACATACCCAATAGATTCTTGATTAGGGGAAGGGA
AAATGAGCAAGATACAGTCCAGTATTCTAAAAACAATCAGCCTTAGGGGATCATTTCAAAAGCATCTGTTTTGGACT
TAAGTCTTTGATACTTAACCAAATTGACTACACAGTGAAAAATTCTAGTGCCTGGGTTTTATAGGGTAGAAGAAAGA
CATGCAGTCAAGTGGCCAATACTTCATGTGAAGATAAGCAATGAGATCCTTCTTGCTGTCTTTCTTTTGACTGTTCT
GGGCAATATCAAATTAGTTTCAGTGGCTTGATTCTAGGCCAAGATTCTGGCAACAGATTGTAGTCTTACCTTGTTTT
CTTCAATCTCACTGGATCTCTCTCTCTTTTTACCCCCCTTAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGT
CCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGCAGGAATGGC
TTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTGAAGTTGGAT
TACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATT
ATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATTTTTGAATGT
TATTATATATTACCTGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAATAAAAACGTTAGAACCTTCTGC
TTAGAGTACATTTAAAAAA

<210> SEQ ID NO 84

<211> Length : 1,956

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 84

>T99080_PEA_4_T2

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAG
GGGTGTTTTTCCGTAAGCATACTCAGCTATTTACCATATAAGATAACTCTTAAGAACTGGAGATAGTCAGCTCCCCT
GGGTTAATTTGAAGCAGAAGAGGGCAGTTGTTATACTGCCCTGTCAGTTGGATGCGGAGTCTTACTCAAAATTCATT
CTCAGCATTCTTCTTTTATGGTATCTTCTTTGGCACTTAGCAGCGCATCAGGTAGGCATCTTCTATTTTTCTTCATT
CCTTAATTTCCTTTGTATCCCTCAAATGGTTATTTATTTGGCTGGAGTCTGTTTTGTTCATTAAGCAAACATGTCTT
TGCTTTGAACATGTCTTTGATTTGATGGATACTTAAATTCCTCATCAAACATTTTGTTGCTATGCATAACGTTTTCT
TTGGCCAACTCCAGCAATTTCCCACATTTTGACATGCAATCATGTTAACTCCCATTTTCTTTTGTAATCCAACATCT
TCTATTTAGATAATTACTTTAACAATCAATGACTTAATATTCTAATCATAAATTTATACAAAAATAAAATTACCTCC

AAAACATTGCTACCTTTCCTAAACATTCAGTCTTGCCACAGTTTAATAAAAGGAAAGAACATTAAAAAGGATAAGAC
ACTGTAATGATTAGATGCTTTTTATAAGCCTAAAGGCATTGTGATTATTTAGACAGAAGAGAAGAAAGTGAAGTGAA
AACCTGATAGTTATGTAGTCTCATGGTTTGCTGTTGAGAGGCTGAACACCAGCTGCTTTCCTTTTCTAGGAAGATAA
TAAAGTGGGCTTTGGCTACAACATAAAGATGTTGGGTTAGACAGTTTCACTACAGTAAGAACAACGGGATGAGTTGC
CCAGGAAATTGTGAAATACTTTCTAATGATCTTTAAAGATATAATGAACACTAATTCATCTGGATTTGTTTAGGTGT
GGTCCTGGTTAAAGGCAAAGGGAAGGATCAGATAACTTCATGTTTTTTCCATTTAACATACCCAATAGATTCTTGAT
TAGGGGAAGGGAAAATGAGCAAGATACAGTCCAGTATTCTAAAAACAATCAGCCTTAGGGGATCATTTCAAAAGCAT
CTGTTTTGGACTTAAGTCTTTGATACTTAACCAAATTGACTACACAGTGAAAAATTCTAGTGCCTGGGTTTTATAGG
GTAGAAGAAAGACATGCAGTCAAGTGGCCAATACTTCATGTGAAGATAAGCAATGAGATCCTTCTTGCTGTCTTTCT
TTTGACTGTTCTGGGCAATATCAAATTAGTTTCAGTGGCTTGATTCTAGGCCAAGATTCTGGCAACAGATTGTAGTC
TTACCTTGTTTTCTTCAATCTCACTGGATCTCTCTCTCTTTTTACCCCCCTTAGGCTGAGGGTAAAAAGCTGGGATT
GGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATA
TGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAAAAAGTCATC
TTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGT
TTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTTTAATGTCAG
ATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAATAAAAACGTT
AGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 85

<211> Length : 1,804

<212> Type : DNA

<213> ORGANISM : Homo sapiens

<400> sequence : 85

>T99080_PEA_4_T4

TTTTGAGACAGAGTCTTGCTCTGTTGCCTAGGCTAAAGTGCAGTAGTGCGATCTCGGCTCACTGCAACCTCCACTTC
CTGGGTTAATTTGAAGCAGAAGAGGGCAGTTGTTATACTGCCCTGTCAGTTGGATGCGGAGTCTTACTCAAAATTCA
TTCTCAGCATTCTTCTTTTATGGTATCTTCTTTGGCACTTAGCAGCGCATCAGGTAGGCATCTTCTATTTTTCTTCA
TTCCTTAATTTCCTTTGTATCCCTCAAATGGTTATTTATTTGGCTGGAGTCTGTTTTGTTCATTAAGCAAACATGTC
TTTGCTTTGAACATGTCTTTGATTTGATGGATACTTAAATTCCTCATCAAACATTTTGTTGCTATGCATAACGTTTT
CTTTGGCCAACTCCAGCAATTTCCCACATTTTGACATGCAATCATGTTAACTCCCATTTTCTTTTGTAATCCAACAT
CTTCTATTTAGATAATTACTTTAACAATCAATGACTTAATATTCTAATCATAAATTTATACAAAAATAAAATTACCT
CCAAAACATTGCTACCTTTCCTAAACATTCAGTCTTGCCACAGTTTAATAAAAGGAAAGAACATTAAAAAGGATAAG
ACACTGTAATGATTAGATGCTTTTTATAAGCCTAAAGGCATTGTGATTATTTAGACAGAAGAGAAGAAAGTGAAGTG
AAAACCTGATAGTTATGTAGTCTCATGGTTTGCTGTTGAGAGGCTGAACACCAGCTGCTTTCCTTTTCTAGGAAGAT
AATAAAGTGGGCTTTGGCTACAACATAAAGATGTTGGGTTAGACAGTTTCACTACAGTAAGAACAACGGGATGAGTT

GCCCAGGAAATTGTGAAATACTTTCTAATGATCTTTAAAGATATAATGAACACTAATTCATCTGGATTTGTTTAGGT
GTGGTCCTGGTTAAAGGCAAAGGGAAGGATCAGATAACTTCATGTTTTTTCCATTTAACATACCCAATAGATTCTTG
ATTAGGGGAAGGGAAAATGAGCAAGATACAGTCCAGTATTCTAAAAACAATCAGCCTTAGGGGATCATTTCAAAAGC
ATCTGTTTTGGACTTAAGTCTTTGATACTTAACCAAATTGACTACACAGTGAAAAATTCTAGTGCCTGGGTTTTATA
GGGTAGAAGAAAGACATGCAGTCAAGTGGCCAATACTTCATGTGAAGATAAGCAATGAGATCCTTCTTGCTGTCTTT
CTTTTGACTGTTCTGGGCAATATCAAATTAGTTTCAGTGGCTTGATTCTAGGCCAAGATTCTGGCAACAGATTGTAG
TCTTACCTTGTTTTCTTCAATCTCACTGGATCTCTCTCTCTTTTTACCCCCCTTAGGCTGAGGGTAAAAAGCTGGGA
TTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCA
TATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAAAAAGTCA
TCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTG
GTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTTTAATGTC
AGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAATAAAAACG
TTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 86

<211> Length : 838

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 86

>T99080_PEA_4_T6

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAA
GGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACA
CATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAAT
GGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTT
AATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGG
ATTACCACTGTACACAAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 87

<211> Length : 606

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 87

>T99080_PEA_4_T9

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTTCCAGTTAGCTTTGCT
ACCAGTTATGCAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAG
GACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACAC
ATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATG
GCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTA
ATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGA
TTACCACTGTACACAAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 88

<211> Length : 698

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 88

>T99080_PEA_4_T10

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAG
GGGTGTTTTTCCGTAAGCATACTCAGGAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTT
CCAGTTAGCTTTGCTACCAGTTATGCAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACC
GGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGA
AGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCA
AATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAG
AACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACC
TGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTA
AAAAA

<210> SEQ ID NO 89

<211> Length : 733

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 89

>T99080_PEA_4_T11

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAG
GGGTGTTTTTCCGTAAGCATACTCAGCTTTGTTTTTGGGAGGCTGAGGAAGGAGGATCATTTGAGCCTGGGAGGTTA
AGGCTGCAATAAGCTGTGACTGTGCCACCATCCTTCAGAAAAAAAAAAGAAAAAGGAAAAGAGGCTGAGGGTAAAAA
GCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGG
TGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAA
AAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAA
CTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTT
TAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAAT
AAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 90

<211> Length : 746

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 90

>T99080_PEA_4_T13

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
CTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGT
CCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAA
CTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGT
TTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGT
CAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACAC
AAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 91

<211> Length : 782

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 91

>T99080_PEA_4_T14

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAG
GGGTGTTTTTCCGTAAGCATACTCAGGAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTT
CCAGTTAGCTTTGCTACCAGTTATGCAGGTATTAATGAATTTAAAAGGATTTAAATCAAGGAATGTTCTCCAACTAC
AGTGGAACTAAACCACATTAAAAAAATAAAAAGGATAACTGGAAAATCCCAAAATATTTGGAAACCATATAGCACAC
TTACTTCTAAAATTGTGGTAGAATACATATAACATAGAAATTATTGTTCTAACCATTTTTAAATGTACAATTCAGTG
GTCTTAAGCACATTCACATTGTTCTGTTTATCTACAGAACGCTTTTCATCTTGCAAAACTGAAACTCTGTATTCATT
AAACACTAACTCCCCATTTTCTCCTTCCCCCATGCCCCTGACAATCATAAATCTACATTCTATTAATTCAACTGCTC
TAGTTACCTCATATAAGTGGAATTTTACAGTATTTGTCCTTTTGTGGCTGGCTTATTTCACTTAGCATAATATCCTC
AGGGTTCATCTGTGTTATATCATGAAAGTAAAAACAATTTCCTTTCTTTGTAAGACGGAATAATATCCTGTTGTATG
TGTATACTTTCA

<210> SEQ ID NO 92

<211> Length : 627

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 92

>T99080_PEA_4_T17

TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGGGTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAG
GGGTGTTTTTCCGTAAGCATACTCAGGTATGTGGCCTGCAGGCTTTGGGGTGGTAACCTAGGTTGTGGGCATAGGAA
CAAGGGACGTGTTCTTGACAAATGTGAACTAGAAGCCGCTGGCTATTTGGTAGCTGCATGGCAAAGAGGTAGTTTGT
AGAGCAATGAGTATGAAAATGCTGTGCAATCGGTAAAACATGGTGTAAATGGAAGATCACCCTGCTGTTATTATTAG
TCAGTGGTGCCTGAGCATCTAGATACACATATTAATGAGCTTCTCTCTTCCAAGGGAAATAGAGGGCTTCCCCAGTG
CTCGCCGTTGTGGCATCATCAAACCAAGTCAGGTTTCTTATAGAAAGGCTAACATTGATTGAAGAGCTGGCATTTAG

ATGACCTGATGTCATGTATATAACATATATAAATGCTTCCTCTGCAGCTGCTGCACTTTCTTCAGACCTCTTTCTCC
AAGCTTCCCTC

<210> SEQ ID NO 93

<211> Length : 917

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 93

>T99080_PEA_4_T18

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGGAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTTCCAGTTAGCTTT
GCTACCAGTTATGCAGGCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGC
AAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCA
CACATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATA
ATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTG
TTAATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAA
GGATTACCACTGTACACAAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 94

<211> Length : 952

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 94

>T99080_PEA_4_T19

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG

TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGCTTTGTTTTTGGGAGGCTGAGGAAGGAGGATCATTTGAGCCTGGGAGGTTAAGGCTGCAATAA
GCTGTGACTGTGCCACCATCCTTCAGAAAAAAAAAAGAAAAAGGAAAAGAGGCTGAGGGTAAAAAGCTGGGATTGGT
AGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGGTCCCATCTCCAAGGTGCGTCATATGC
AGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAAACTTCAACAATGAAAAAGTCATCTTG
AAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAGTTTTCTAAGATAAACTCAGTGGTTTG
GTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAGTCAATAAGTTATTTTAATGTCAGATT
TTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACACAAATCTAATCAATAAAAACGTTAGA
ACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 95

<211> Length : 1,001

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 95

>T99080_PEA_4_T20

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGGAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTTCCAGTTAGCTTT
GCTACCAGTTATGCAGGTATTAATGAATTTAAAAGGATTTAAATCAAGGAATGTTCTCCAACTACAGTGGAACTAAA
CCACATTAAAAAAATAAAAAGGATAACTGGAAAATCCCAAAATATTTGGAAACCATATAGCACACTTACTTCTAAAA
TTGTGGTAGAATACATATAACATAGAAATTATTGTTCTAACCATTTTTAAATGTACAATTCAGTGGTCTTAAGCACA
TTCACATTGTTCTGTTTATCTACAGAACGCTTTTCATCTTGCAAAACTGAAACTCTGTATTCATTAAACACTAACTC
CCCATTTTCTCCTTCCCCCATGCCCCTGACAATCATAAATCTACATTCTATTAATTCAACTGCTCTAGTTACCTCAT
ATAAGTGGAATTTTACAGTATTTGTCCTTTTGTGGCTGGCTTATTTCACTTAGCATAATATCCTCAGGGTTCATCTG
TGTTATATCATGAAAGTAAAAACAATTTCCTTTCTTTGTAAGACGGAATAATATCCTGTTGTATGTGTATACTTTCA

<210> SEQ ID NO 96

<211> Length : 846

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 96

>T99080_PEA_4_T21

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAGG
TTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTCC
GTAAGCATACTCAGGTATGTGGCCTGCAGGCTTTGGGGTGGTAACCTAGGTTGTGGGCATAGGAACAAGGGACGTGT
TCTTGACAAATGTGAACTAGAAGCCGCTGGCTATTTGGTAGCTGCATGGCAAAGAGGTAGTTTGTAGAGCAATGAGT
ATGAAAATGCTGTGCAATCGGTAAAACATGGTGTAAATGGAAGATCACCCTGCTGTTATTATTAGTCAGTGGTGCCT
GAGCATCTAGATACACATATTAATGAGCTTCTCTCTTCCAAGGGAAATAGAGGGCTTCCCCAGTGCTCGCCGTTGTG
GCATCATCAAACCAAGTCAGGTTTCTTATAGAAAGGCTAACATTGATTGAAGAGCTGGCATTTAGATGACCTGATGT
CATGTATATAACATATATAAATGCTTCCTCTGCAGCTGCTGCACTTTCTTCAGACCTCTTTCTCCAAGCTTCCCTC

<210> SEQ ID NO 97

<211> Length : 4,539

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 97

>T08446_PEA_1_T2

GCTACGGAAGGGTTTTCAGCAGGGAGAGATGGGACCAGCATGCTCTCCATCCTGTTGCCCCGCCTCACGTCTGGCCC
CTGCTTCTCCAGTCCCCCACCCCAGACCACACACGAAGAAGCAGTCCTGTCCTCAGCCCAGCCCTCACCTCCCCCGA
CCTGCCATCCTGCTTCATGCTCAGGGCGGTGTGTGGAGCGCCCGGGGCTCTGGACCCGCGCTGCCAGATAACAATGC
TCTCGTTGTCTCTTTGCTCCCATCTCTGGGGGCCTCTGATTCTTTCTGCTCTACAGGCACGCAGCACTGACAGCCTG
GATGGCCCAGGGGAGGGCTCGGTGCAGCCTCTACCCACTGCTGGGGGGCCCAGTGTGAAGGGGAAGCCTGGGAAGAG
GCTCTCAGCTCCTCGAGGCCCCTTCCCGCGGCTGGCTGACTGCGCCCATTTCCACTACGAGAACGTTGACTTTGGCC
ACATTCAGCTCCTGCTGTCTCCAGACCGTGAAGGGCCCAGCCTCTCTGGAGAGAATGAGCTGGTGTTCGGGGTGCAG
GTGACCTGTCAGGGCCGTTCCTGGCCGGTTCTCCGGAGTTACGATGACTTTCGTTCCCTGGATGCCCACCTCCACCG
GTGCATATTTGACCGGAGGTTCTCCTGCCTTCCGGAGCTTCCCCCGCCCCCCGAGGGTGCCAGGGCTGCCCAGATGC
TGGTGCCACTGCTGCTGCAGTACCTGGAGACACTGTCAGGACTGGTGGACAGTAACCTCAACTGCGGGCCTGTGCTC
ACCTGGATGGAGCTGGACAATCACGGCCGGCGACTGCTCCTCAGTGAGGAGGCGTCACTCAATATCCCTGCAGTGGC

GGCCGCCCATGTGATCAAACGGTATACAGCCCAGGCGCCAGATGAGCTGTCCTTTGAGGTGGGAGACATTGTCTCGG
TGATCGACATGCCACCCACAGAGGATCGGAGCTGGTGGCGGGGCAAGCGAGGCTTCCAGGTCGGGTTCTTCCCCAGT
GAGTGTGTGGAACTCTTCACAGAGCGGCCAGGTCCGGGCCTGAAGGCGGATGCCGATGGCCCCCCATGTGGCATCCC
GGCTCCCCAGGGTATCTCGTCTCTGACCTCAGCTGTGCCACGGCCTCGTGGGAAGCTGGCCGGCCTGCTCCGCACCT
TCATGCGCTCCCGCCCTTCTCGGCAGCGGCTGCGGCAGCGGGGAATCCTGCGACAGAGGGTGTTTGGCTGCGATCTT
GGCGAGCACCTCAGCAACTCAGGCCAGGATGTGCCCCAGGTGCTGCGCTGCTGCTCCGAGTTCATTGAGGCCCACGG
GGTGGTGGATGGGATCTACCGGCTCTCAGGCGTGTCTTCCAACATCCAGAGGCTTCGGCACGAGTTTGACAGTGAGA
GGATCCCGGAGCTGTCTGGCCCTGCATTCCTGCAGGACATCCACAGCGTGTCCTCCCTCTGCAAGCTCTACTTCCGA
GAGCTTCCGAACCCTCTGCTCACCTACCAGCTCTATGGGAAGTTCAGTGAGGCCATGTCAGTGCCTGGGGAGGAGGA
GCGTCTGGTGCGGGTGCACGATGTCATCCAGCAGCTGCCCCCACCACATTACAGGACCCTGGAGTACCTGCTGAGGC
ACCTGGCCCGCATGGCGAGACACAGTGCCAACACCAGCATGCATGCCCGCAACCTGGCCATTGTCTGGGCACCCAAC
CTGCTACGGTCCATGGAGCTGGAGTCAGTGGGAATGGGTGGCGCGGCGGCGTTCCGGGAAGTTCGGGTGCAGTCGGT
GGTGGTGGAGTTTCTGCTCACCCATGTGGACGTCCTGTTCAGCGACACCTTCACCTCCGCCGGCCTCGACCCTGCAG
GCCGCTGCCTGCTCCCCAGGCCCAAGTCCCTTGCGGGCAGCTGCCCCTCCACCCGCCTGCTGACGCTGGAGGAAGCC
CAGGCACGCACCCAGGGCCGGCTGGGGACGCCCACGGAGCCCACAACTCCCAAGGCCCCGGCCTCACCTGCGGAAAG
GAGGAAAGGGGAGAGAGGGGAGAAGCAGCGGAAGCCAGGGGGCAGCAGCTGGAAGACGTTCTTTGCACTGGGCCGGG
GCCCCAGTGTCCCTCGAAAGAAGCCCCTGCCCTGGCTGGGGGGCACCCGTGCCCCACCGCAGCCTTCAGGCAGCAGA
CCCGACACCGTCACACTGAGATCTGCCAAGAGCGAGGAGTCTCTGTCATCGCAGGCCAGCGGGGCTGGCCTCCAGAG
GCTGCACAGGCTGCGGCGACCCCACTCCAGCAGCGACGCTTTCCCTGTGGGCCCAGCACCTGCTGGCTCCTGCGAGA
GCCTGTCCTCGTCCTCCTCCTCCGAGTCCTCCTCCTCTGAGTCCTCCTCTTCCTCCTCTGAGTCCTCAGCAGCTGGG
CTGGGGGCACTCTCTGGGTCTCCCTCACACCGTACCTCAGCCTGGCTAGATGATGGTGATGAGCTGGACTTCAGCCC
ACCCCGCTGCCTGGAGGGACTCCGGGGGCTGGACTTTGATCCCTTAACCTTCCGCTGCAGCAGCCCCACCCCAGGGG
ATCCCGCACCTCCCGCCAGCCCAGCACCCCCCGCCCCTGCCTCTGCCTTCCCACCCAGGGTGACCCCCCAGGCCATC
TCGCCCCGGGGGCCCACCAGCCCCGCCTCGCCTGCTGCCCTAGACATCTCAGAGCCCCTGGCTGTATCAGTGCCACC
CGCTGTCCTAGAACTGCTGGGGGCTGGGGGAGCACCTGCCTCAGCCACCCCAACACCAGCTCTCAGCCCCGGCCGGA
GCCTGCGCCCCCATCTCATACCCCTGCTGCTGCGAGGAGCCGAGGCCCCGCTGACTGACGCCTGCCAGCAGGAGATG
TGCAGCAAGCTCCGGGGAGCCCAGGGCCCACTCGGTCCTGATATGGAGTCACCACTGCCACCCCCTCCCCTGTCTCT
CCTGCGCCCTGGGGGTGCCCCACCCCCGCCCCCTAAGAACCCAGCACGCCTCATGGCCCTGGCCCTGGCTGAGCGGG
CTCAGCAGGTGGCCGAGCAACAGAGCCAGCAGGAGTGTGGGGGCACCCCACCTGCTTCCCAATCCCCCTTCCACCGC
TCGCTGTCTCTGGAGGTGGGCGGGGAGCCCCTGGGGACCTCAGGGAGTGGGCCACCTCCCAACTCCCTAGCACACCC
GGGTGCCTGGGTCCCGGGACCCCCACCCTACTTACCAAGGCAACAAAGTGATGGGAGCCTGCTGAGGAGCCAGCGGC
CCATGGGGACCTCAAGGAGGGGACTCCGAGGCCCTGCCCAGGTCAGTGCCCAGCTCAGGGCAGGTGGCGGGGGCAGG
GATGCGCCAGAGGCAGCAGCCCAGTCCCCATGTTCTGTCCCCTCACAGGTTCCTACCCCCGGCTTCTTCTCCCCAGC
CCCCAGGGAGTGCCTGCCACCCTTCCTCGGGGTCCCCAAGCCAGGCTTGTACCCCCTGGGCCCCCCATCCTTCCAGC
CCAGTTCCCCAGCCCCAGTCTGGAGGAGCTCTCTGGGCCCCCCTGCACCACTCGACAGGGGAGAGAACCTGTACTAT
GAGATCGGGGCAAGTGAGGGGTCCCCCTATTCTGGCCCCACCCGCTCCTGGAGTCCCTTTCGCTCCATGCCCCCCGA
CAGGCTCAATGCCTCCTACGGCATGCTTGGCCAATCACCCCCACTCCACAGGTCCCCCGACTTCCTGCTCAGCTACC
CGCCAGCCCCCTCCTGCTTTCCCCCTGACCACCTTGGCTACTCAGCCCCCCAGCACCCTGCTCGGCGCCCTACACCG

CCTGAGCCCCTCTACGTCAACCTAGCTCTAGGGCCCAGGGGTCCCTCACCTGCCTCTTCCTCCTCCTCTTCCCCTCC
TGCCCACCCCCGAAGCCGTTCAGATCCCGGTCCCCCAGTCCCCCGCCTTCCCCAGAAACAACGGGCACCCTGGGGAC
CCCGTACCCCTCATAGGGTGCCGGGTCCCTGGGGCCCTCCTGAGCCTCTCCTGCTCTACAGGGCAGCCCCGCCAGCC
TACGGAAGGGGGGGCGAGCTCCACCGAGGGTCCTTGTACAGAAATGGAGGGCAAAGAGGGGAGGGGGCTGGTCCCCC
ACCCCCTTACCCCACTCCCAGCTGGTCCCTCCACTCTGAGGGCCAGACCCGAAGCTACTGCTGAGCACCAGCTGGGA
GGGGCCGTCCTTCCTTCCCTTCACCCTCACTGGATCTTGGCCCAACCAAATCCCTTGTTTTGTATTTTCTTGAACCC
CGACCACTACCCCAGGTTTCTAACTTTGTAACTTGCTTCTGATGTGGGTCCCTAACCTATAATCTCAGCTTCCCTAC
CCTGGACTGAAGGGTCTGCCCATCCCCCCACCACCCTCCATCCTGGGGGCCCTCGCACAAATCTGGGGTGGGAGGGG
CTAGGCTGACCCCATCCTCCTCTCCCTCCAGGAGCCCCCAGCATGTCCTGACCTGTGCACGGGGATGGGGGGACAAC
TCCTACCCTTCTTTCCCCACATGCCCCACTAAACCATCTGACAACATTAATGAATAAAATGGTGAAAATGTGA

<210> SEQ ID NO 98

<211> Length : 968

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 98

>T08446_PEA_1_T22

GCTACGGAAGGGTTTTCAGCAGGGAGAGATGGGACCAGCATGCTCTCCATCCTGTTGCCCCGCCTCACGTCTGGCCC
CTGCTTCTCCAGTCCCCCACCCCAGACCACACACGAAGAAGCAGTCCTGTCCTCAGCCCAGCCCTCACCTCCCCCGA
CCTGCCATCCTGCTTCATGCTCAGGGCGGTGTGTGGAGCGCCCGGGGCTCTGGACCCGCGCTGCCAGATAACAATGC
TCTCGTTGTCTCTTTGCTCCCATCTCTGGGGGCCTCTGATTCTTTCTGCTCTACAGGCACGCAGCACTGACAGCCTG
GATGGCCCAGGGGAGGGCTCGGTGCAGCCTCTACCCACTGCTGGGGGGCCCAGTGTGAAGGGGAAGCCTGGGAAGAG
GCTCTCAGCTCCTCGAGGCCCCTTCCCGCGGCTGGCTGACTGCGCCCATTTCCACTACGAGAACGTTGACTTTGGCC
ACATTCAGCTCCTGCTGTCTCCAGACCGTGAAGGGCCCAGCCTCTCTGGAGAGAATGAGCTGGTGTTCGGGGTGCAG
GTGACCTGTCAGGGCCGTTCCTGGCCGGTTCTCCGGAGTTACGATGACTTTCGTTCCCTGGATGCCCACCTCCACCG
GTGCATATTTGACCGGAGGTTCTCCTGCCTTCCGGAGCTTCCCCCGCCCCCCGAGGGTGCCAGGGCTGCCCAGATGC
TGGTGCCACTGCTGCTGCAGTACCTGGAGACACTGTCAGGACTGGTGGACAGTAACCTCAACTGCGGGCCTGTGCTC
ACCTGGATGGAGGTGGGCCTGGGCAGGGGGCTTGGAGATTCCGAGTGGGTGAGGGGGTGCGTGTGCCACCACGCCCA
GCACAGAGAGATTTTAGATGGCAACCGTGTGGCATCTGCTGTGGAGGATGAAGGTGCAGAGGTGGATGGGGAAGCCT
TCAGGTGGGGAAGCCTTTGGGTGGGAGAGTCCTGGGACATGTGA

<210> SEQ ID NO 99

<211> Length : 3,615

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 99

>HUMCA1XIA_T16

ACACAGTACTCTCAGCTTGTTGGTGGAAGCCCCTCATCTGCCTTCATTCTGAAGGCAGGGCCCGGCAGAGGAAGGAT
CAGAGGGTCGCGGCCGGAGGGTCCCGGCCGGTGGGGCCAACTCAGAGGGAGAGGAAAGGGCTAGAGACACGAAGAAC
GCAAACCATCAAATTTAGAAGAAAAAGCCCTTTGACTTTTTCCCCCTCTCCCTCCCCAATGGCTGTGTAGCAAACAT
CCCTGGCGATACCTTGGAAAGGACGAAGTTGGTCTGCAGTCGCAATTTCGTGGGTTGAGTTCACAGTTGTGAGTGCG
GGGCTCGGAGATGGAGCCGTGGTCCTCTAGGTGGAAAACGAAACGGTGGCTCTGGGATTTCACCGTAACAACCCTCG
CATTGACCTTCCTCTTCCAAGCTAGAGAGGTCAGAGGAGCTGCTCCAGTTGATGTACTAAAAGCACTAGATTTTCAC
AATTCTCCAGAGGGAATATCAAAAACAACGGGATTTTGCACAAACAGAAAGAATTCTAAAGGCTCAGATACTGCTTA
CAGAGTTTCAAAGCAAGCACAACTCAGTGCCCCAACAAAACAGTTATTTCCAGGTGGAACTTTCCCAGAAGACTTTT
CAATACTATTTACAGTAAAACCAAAAAAAGGAATTCAGTCTTTCCTTTTATCTATATATAATGAGCATGGTATTCAG
CAAATTGGTGTTGAGGTTGGGAGATCACCTGTTTTTCTGTTTGAAGACCACACTGGAAAACCTGCCCCAGAAGACTA
TCCCCTCTTCAGAACTGTTAACATCGCTGACGGGAAGTGGCATCGGGTAGCAATCAGCGTGGAGAAGAAAACTGTGA
CAATGATTGTTGATTGTAAGAAGAAAACCACGAAACCACTTGATAGAAGTGAGAGAGCAATTGTTGATACCAATGGA
ATCACGGTTTTTGGAACAAGGATTTTGGATGAAGAAGTTTTTGAGGGGGACATTCAGCAGTTTTTGATCACAGGTGA
TCCCAAGGCAGCATATGACTACTGTGAGCATTATAGTCCAGACTGTGACTCTTCAGCACCCAAGGCTGCTCAAGCTC
AGGAACCTCAGATAGATGAGTATGCACCAGAGGATATAATCGAATATGACTATGAGTATGGGGAAGCAGAGTATAAA
GAGGCTGAAAGTGTAACAGAGGGACCCACTGTAACTGAGGAGACAATAGCACAGACGGAGGCAAACATCGTTGATGA
TTTTCAAGAATACAACTATGGAACAATGGAAAGTTACCAGACAGAAGCTCCTAGGCATGTTTCTGGGACAAATGAGC
CAAATCCAGTTGAAGAAATATTTACTGAAGAATATCTAACGGGAGAGGATTATGATTCCCAGAGGAAAAATTCTGAG
GATACACTATATGAAAACAAAGAAATAGACGGCAGGGATTCTGATCTTCTGGTAGATGGAGATTTAGGCGAATATGA
TTTTTATGAATATAAAGAATATGAAGATAAACCAACAAGCCCCCCTAATGAAGAATTTGGTCCAGGTGTACCAGCAG
AAACTGATATTACAGAAACAAGCATAAATGGCCATGGTGCATATGGAGAGAAAGGACAGAAAGGAGAACCAGCAGTG
GTTGAGCCTGGTATGCTTGTCGAAGGACCACCAGGACCAGCAGGACCTGCAGGTATTATGGGTCCTCCAGGTCTACA
AGGCCCCACTGGACCCCCTGGTGACCCTGGCGATAGGGGCCCCCCAGGACGTCCTGGCTTACCAGGGGCTGATGGTC
TACCTGGTCCTCCTGGTACTATGTTGATGTTACCGTTCCGTTATGGTGGTGATGGTTCCAAAGGACCAACCATCTCT
GCTCAGGAAGCTCAGGCTCAAGCTATTCTTCAGCAGGCTCGGATTGCTCTGAGAGGCCCACCTGGCCCAATGGGTCT
AACTGGAAGACCAGGTCCTGTGGGGGGGCCTGGTTCATCTGGGGCCAAAGGTGAGAGTGGTGATCCAGGTCCTCAGG
GCCCTCGAGGCGTCCAGGGTCCCCCTGGTCCAACGGGAAAACCTGGAAAAAGGGGTCGTCCAGGTGCAGATGGAGGA
AGAGGAATGCCAGGAGAACCTGGGGCAAAGGGAGATCGAGGGTTTGATGGACTTCCGGGTCTGCCAGGTGACAAAGG
TCACAGGGGTGAACGAGGTCCTCAAGGTCCTCCAGGTCCTCCTGGTGATGATGGAATGAGGGGAGAAGATGGAGAAA
TTGGACCAAGAGGTCTTCCAGGTGAAGCTGGCCCACGAGGTTTGCTGGGTCCAAGGGGAACTCCAGGAGCTCCAGGG
CAGCCTGGTATGGCAGGTGTAGATGGCCCCCCAGGACCAAAAGGGAACATGGGTCCCCAAGGGGAGCCTGGGCCTCC
AGGTCAACAAGGGAATCCAGGACCTCAGGGTCTTCCTGGTCCACAAGGTCCAATTGGTCCTCCTGGTGAAAAAGGAC

CACAAGGAAAACCAGGACTTGCTGGACTTCCTGGTGCTGATGGGCCTCCTGGTCATCCTGGGAAAGAAGGCCAGTCT
GGAGAAAAGGGGGCTCTGGGTCCCCCTGGTCCACAAGGTCCTATTGGATACCCGGGCCCCCGGGGAGTAAAGGGAGC
AGATGGTGTCAGAGGTCTCAAGGGATCTAAAGGTGAAAAGGGTGAAGATGGTTTTCCAGGATTCAAAGGTGACATGG
GTCTAAAAGGTGACAGAGGAGAAGTTGGTCAAATTGGCCCAAGAGGGGAAGATGGCCCTGAAGGACCCAAAGGTCGA
GCAGGCCCAACTGGAGACCCAGGTCCTTCAGGTCAAGCAGGAGAAAAGGGAAAACTTGGAGTTCCAGGATTACCAGG
ATATCCAGGAAGACAAGGTCCAAAGGGTTCCACTGGATTCCCTGGGTTTCCAGGTGCCAATGGAGAGAAAGGTGCAC
GGGGAGTAGCTGGCAAACCAGGCCCTCGGGGTCAGCGTGGTCCAACGGGTCCTCGAGGTTCAAGAGGTGCAAGAGGT
CCCACTGGGAAACCTGGGCCAAAGGGCACTTCAGGTGGCGATGGCCCTCCTGGCCCTCCAGGTGAAAGAGGTCCTCA
AGGACCTCAGGGTCCAGTTGGATTCCCTGGACCAAAAGGCCCTCCTGGACCACCTGGGAAGGATGGGCTGCCAGGAC
ACCCTGGGCAACGTGGGGAGACTGGATTTCAAGGCAAGACCGGCCCTCCTGGGCCAGGGGGAGTGGTTGGACCACAG
GGACCAACCGGTGAGACTGGTCCAATAGGGGAACGTGGGCATCCTGGCCCTCCTGGCCCTCCTGGTGAGCAAGGTCT
TCCTGGTGCTGCAGGAAAAGAAGGTGCAAAGGGTGATCCAGGTCCTCAAGGTATCTCAGGGAAAGATGGACCAGCAG
GATTACGTGGTTTCCCAGGGGAAAGAGGTCTTCCTGGAGCTCAGGGTGCACCTGGACTGAAAGGAGGGGAAGGTCCC
CAGGGCCCACCAGGTCCAGTTGTAAGTATGATGATAATAAATAGCCAGACAATCATGGTTGTGAATTACAGCTCTTC
TTTCATTACTCTCATGCTGTGATTCCACAGTGTGTGGGAGAGAAAATAAACACATGTCAATCAAATCAGTCAA

<210> SEQ ID NO 100

<211> Length : 2,648

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 100

>HUMCA1XIA_T17

ACACAGTACTCTCAGCTTGTTGGTGGAAGCCCCTCATCTGCCTTCATTCTGAAGGCAGGGCCCGGCAGAGGAAGGAT
CAGAGGGTCGCGGCCGGAGGGTCCCGGCCGGTGGGGCCAACTCAGAGGGAGAGGAAAGGGCTAGAGACACGAAGAAC
GCAAACCATCAAATTTAGAAGAAAAAGCCCTTTGACTTTTTCCCCCTCTCCCTCCCCAATGGCTGTGTAGCAAACAT
CCCTGGCGATACCTTGGAAAGGACGAAGTTGGTCTGCAGTCGCAATTTCGTGGGTTGAGTTCACAGTTGTGAGTGCG
GGGCTCGGAGATGGAGCCGTGGTCCTCTAGGTGGAAAACGAAACGGTGGCTCTGGGATTTCACCGTAACAACCCTCG
CATTGACCTTCCTCTTCCAAGCTAGAGAGGTCAGAGGAGCTGCTCCAGTTGATGTACTAAAAGCACTAGATTTTCAC
AATTCTCCAGAGGGAATATCAAAAACAACGGGATTTTGCACAAACAGAAAGAATTCTAAAGGCTCAGATACTGCTTA
CAGAGTTTCAAAGCAAGCACAACTCAGTGCCCCAACAAAACAGTTATTTCCAGGTGGAACTTTCCCAGAAGACTTTT
CAATACTATTTACAGTAAAACCAAAAAAAGGAATTCAGTCTTTCCTTTTATCTATATATAATGAGCATGGTATTCAG
CAAATTGGTGTTGAGGTTGGGAGATCACCTGTTTTTCTGTTTGAAGACCACACTGGAAAACCTGCCCCAGAAGACTA
TCCCCTCTTCAGAACTGTTAACATCGCTGACGGGAAGTGGCATCGGGTAGCAATCAGCGTGGAGAAGAAAACTGTGA
CAATGATTGTTGATTGTAAGAAGAAAACCACGAAACCACTTGATAGAAGTGAGAGAGCAATTGTTGATACCAATGGA
ATCACGGTTTTTGGAACAAGGATTTTGGATGAAGAAGTTTTTGAGGGGGACATTCAGCAGTTTTTGATCACAGGTGA

TCCCAAGGCAGCATATGACTACTGTGAGCATTATAGTCCAGACTGTGACTCTTCAGCACCCAAGGCTGCTCAAGCTC
AGGAACCTCAGATAGATGAGTATGCACCAGAGGATATAATCGAATATGACTATGAGTATGGGGAAGCAGAGTATAAA
GAGGCTGAAAGTGTAACAGAGGGACCCACTGTAACTGAGGAGACAATAGCACAGACGGAGGCAAACATCGTTGATGA
TTTTCAAGAATACAACTATGGAACAATGGAAAGTTACCAGACAGAAGCTCCTAGGCATGTTTCTGGGACAAATGAGC
CAAATCCAGTTGAAGAAATATTTACTGAAGAATATCTAACGGGAGAGGATTATGATTCCCAGAGGAAAAATTCTGAG
GATACACTATATGAAAACAAAGAAATAGACGGCAGGGATTCTGATCTTCTGGTAGATGGAGATTTAGGCGAATATGA
TTTTTATGAATATAAAGAATATGAAGATAAACCAACAAGCCCCCCTAATGAAGAATTTGGTCCAGGTGTACCAGCAG
AAACTGATATTACAGAAACAAGCATAAATGGCCATGGTGCATATGGAGAGAAAGGACAGAAAGGAGAACCAGCAGTG
GTTGAGCCTGGTATGCTTGTCGAAGGACCACCAGGACCAGCAGGACCTGCAGGTATTATGGGTCCTCCAGGTCTACA
AGGCCCCACTGGACCCCCTGGTGACCCTGGCGATAGGGGCCCCCCAGGACGTCCTGGCTTACCAGGGGCTGATGGTC
TACCTGGTCCTCCTGGTACTATGTTGATGTTACCGTTCCGTTATGGTGGTGATGGTTCCAAAGGACCAACCATCTCT
GCTCAGGAAGCTCAGGCTCAAGCTATTCTTCAGCAGGCTCGGATTGCTCTGAGAGGCCCACCTGGCCCAATGGGTCT
AACTGGAAGACCAGGTCCTGTGGGGGGGCCTGGTTCATCTGGGGCCAAAGGTGAGAGTGGTGATCCAGGTCCTCAGG
GCCCTCGAGGCGTCCAGGGTCCCCCTGGTCCAACGGGAAAACCTGGAAAAAGGGGTCGTCCAGGTGCAGATGGAGGA
AGAGGAATGCCAGGAGAACCTGGGGCAAAGGGAGATCGAGGGTTTGATGGACTTCCGGGTCTGCCAGGTGACAAAGG
TCACAGGGGTGAACGAGGTCCTCAAGGTCCTCCAGGTCCTCCTGGTGATGATGGAATGAGGGGAGAAGATGGAGAAA
TTGGACCAAGAGGTCTTCCAGGTGAAGCTGGCCCACGAGGTTTGCTGGGTCCAAGGGGAACTCCAGGAGCTCCAGGG
CAGCCTGGTATGGCAGGTGTAGATGGCCCCCCAGGACCAAAAGGGAACATGGGTCCCCAAGGGGAGCCTGGGCCTCC
AGGTCAACAAGGGAATCCAGGACCTCAGGGTCTTCCTGGTCCACAAGGTCCAATTGGTCCTCCTGGTGAAAAAATGT
GCTGCAATCTAAGTTTCGGAATACTTATACCACTCCAGAAATAATCCTCGTGTTAATATTCAGTCCATGTTTCCACC
CCCTCAGACCTAGCTAACAATTAATTTGCTTTCTGTCTCTGTAGATTTGAGTTTTCTGAATATTTCATGTGAATGAA
ACTTATGAATAAATTTATGATTTAATCATA

<210> SEQ ID NO 101

<211> Length : 3,475

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 101

>HUMCA1XIA_T19

ACACAGTACTCTCAGCTTGTTGGTGGAAGCCCCTCATCTGCCTTCATTCTGAAGGCAGGGCCCGGCAGAGGAAGGAT
CAGAGGGTCGCGGCCGGAGGGTCCCGGCCGGTGGGGCCAACTCAGAGGGAGAGGAAAGGGCTAGAGACACGAAGAAC
GCAAACCATCAAATTTAGAAGAAAAAGCCCTTTGACTTTTTCCCCCTCTCCCTCCCCAATGGCTGTGTAGCAAACAT
CCCTGGCGATACCTTGGAAAGGACGAAGTTGGTCTGCAGTCGCAATTTCGTGGGTTGAGTTCACAGTTGTGAGTGCG
GGGCTCGGAGATGGAGCCGTGGTCCTCTAGGTGGAAAACGAAACGGTGGCTCTGGGATTTCACCGTAACAACCCTCG
CATTGACCTTCCTCTTCCAAGCTAGAGAGGTCAGAGGAGCTGCTCCAGTTGATGTACTAAAAGCACTAGATTTTCAC

AATTCTCCAGAGGGAATATCAAAAACAACGGGATTTTGCACAAACAGAAAGAATTCTAAAGGCTCAGATACTGCTTA
CAGAGTTTCAAAGCAAGCACAACTCAGTGCCCCAACAAAACAGTTATTTCCAGGTGGAACTTTCCCAGAAGACTTTT
CAATACTATTTACAGTAAAACCAAAAAAAGGAATTCAGTCTTTCCTTTTATCTATATATAATGAGCATGGTATTCAG
CAAATTGGTGTTGAGGTTGGGAGATCACCTGTTTTTCTGTTTGAAGACCACACTGGAAAACCTGCCCCAGAAGACTA
TCCCCTCTTCAGAACTGTTAACATCGCTGACGGGAAGTGGCATCGGGTAGCAATCAGCGTGGAGAAGAAAACTGTGA
CAATGATTGTTGATTGTAAGAAGAAAACCACGAAACCACTTGATAGAAGTGAGAGAGCAATTGTTGATACCAATGGA
ATCACGGTTTTTGGAACAAGGATTTTGGATGAAGAAGTTTTTGAGGGGGACATTCAGCAGTTTTTGATCACAGGTGA
TCCCAAGGCAGCATATGACTACTGTGAGCATTATAGTCCAGACTGTGACTCTTCAGCACCCAAGGCTGCTCAAGCTC
AGGAACCTCAGATAGATGAGTATGCACCAGAGGATATAATCGAATATGACTATGAGTATGGGGAAGCAGAGTATAAA
GAGGCTGAAAGTGTAACAGAGGGACCCACTGTAACTGAGGAGACAATAGCACAGACGGAGGCAAACATCGTTGATGA
TTTTCAAGAATACAACTATGGAACAATGGAAAGTTACCAGACAGAAGCTCCTAGGCATGTTTCTGGGACAAATGAGC
CAAATCCAGTTGAAGAAATATTTACTGAAGAATATCTAACGGGAGAGGATTATGATTCCCAGAGGAAAAATTCTGAG
GATACACTATATGAAAACAAAGAAATAGACGGCAGGGATTCTGATCTTCTGGTAGATGGAGATTTAGGCGAATATGA
TTTTTATGAATATAAAGAATATGAAGATAAACCAACAAGCCCCCCTAATGAAGAATTTGGTCCAGGTGTACCAGCAG
AAACTGATATTACAGAAACAAGCATAAATGGCCATGGTGCATATGGAGAGAAAGGACAGAAAGGAGAACCAGCAGTG
GTTGAGCCTGGTATGCTTGTCGAAGGACCACCAGGACCAGCAGGACCTGCAGGTATTATGGGTCCTCCAGGTCTACA
AGGCCCCACTGGACCCCCTGGTGACCCTGGCGATAGGGGCCCCCCAGGACGTCCTGGCTTACCAGGGGCTGATGGTC
TACCTGGTCCTCCTGGTACTATGTTGATGTTACCGTTCCGTTATGGTGGTGATGGTTCCAAAGGACCAACCATCTCT
GCTCAGGAAGCTCAGGCTCAAGCTATTCTTCAGCAGGCTCGGATTGCTCTGAGAGGCCCACCTGGCCCAATGGGTCT
AACTGGAAGACCAGGTCCTGTGGGGGGGCCTGGTTCATCTGGGGCCAAAGGTGAGAGTGGTGATCCAGGTCCTCAGG
GCCCTCGAGGCGTCCAGGGTCCCCCTGGTCCAACGGGAAAACCTGGAAAAAGGGGTCGTCCAGGTGCAGATGGAGGA
AGAGGAATGCCAGGAGAACCTGGGGCAAAGGGAGATCGAGGGTTTGATGGACTTCCGGGTCTGCCAGGTGACAAAGG
TCACAGGGGTGAACGAGGTCCTCAAGGTCCTCCAGGTCCTCCTGGTGATGATGGAATGAGGGGAGAAGATGGAGAAA
TTGGACCAAGAGGTCTTCCAGGTGAAGCTGGTATGGCAGGTGTAGATGGCCCCCCAGGACCAAAAGGGAACATGGGT
CCCCAAGGGGAGCCTGGGCCTCCAGGTCAACAAGGGAATCCAGGACCTCAGGGTCTTCCTGGTCCACAAGGTCCAAT
TGGTCCTCCTGGTGAAAAAGTCTCCTTTTCTTTCTCATTATTTTACAAAAAAGTAATTAAGTTTGCTTGTGACAAAA
GATTTGTAGGAAGACATGATGAAAGGAAAGTTGTGAAGTTGTCATTGCCATTATATCTCATATATGAATAGCAAATG
ATGTTCTTGTTAATAACTCAATGTCTGGATCAAAACAAGAATAAATCTATAATTAAACACATGTCTTTTTCCCTGAT
CTCCACTGTGAGATTCCTTGAGTAATATTTTGTCCCCTGTAGTCATAGCACATTTCTATCTGGCTCTTTCCAACACC
TTTTTTCTTTCATTATTTTTGTTGATTTCTACAAACATATTAATTAAAAAAAACTAATAGCTTTATCGAAGTGTAAT
TTAAATACTATAAATATTCCCTTGTTTTAAGTGTACAAGTGAATTATTTTTACTAAATTTACAGATGTGCTGCAATC
TAAGTTTCGGAATACTTATACCACTCCAGAAATAATCCTCGTGTTAATATTCAGTCCATGTTTCCACCCCCTCAGAC
CTAGCTAACAATTAATTTGCTTTCTGTCTCTGTAGATTTGAGTTTTCTGAATATTTCATGTGAATGAAACTTATGAA
TAAATTTATGATTTAATCATATGAATGTGTATGTGACCTTTATGTCTGACTTCTTTCATTTTAGGTTCATCCATGCT
GTAGCATACATATATTAGTACTTTGCTCCTTATTTTGTTCTCATAGTATTCCATTGATTGGGTATACCAGGTTCTGT
TTACTTTTACTTGGCAGTTGATAGAATAGGTGTAGTTTATACTTTTTCGCTATTCTCCATACCGGTGCTGTAGTGAA
TAATTGCATACAAGTCTTTGTATAGATGTGTTTTCATTCTTTTTGGTATATACTTAGAAGCAGAATTCTTGTGTTAT
GGTAAACTTATATTCAATATTTTGTGAATTCCACTCTTTTCCATATCGATTGTACCATTTTCCCTTCCAAGTAACCA

TGTATGAGGATAGTCATTTCTGCACATTCTCACTAATGCTTGTTATTGTCTGTCTTCTTGATTACGATCATTCTCGT
TGGTGTGAAA

<210> SEQ ID NO 102

<211> Length : 1,271

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 102

>HUMCA1XIA_T20

ACACAGTACTCTCAGCTTGTTGGTGGAAGCCCCTCATCTGCCTTCATTCTGAAGGCAGGGCCCGGCAGAGGAAGGAT
CAGAGGGTCGCGGCCGGAGGGTCCCGGCCGGTGGGGCCAACTCAGAGGGAGAGGAAAGGGCTAGAGACACGAAGAAC
GCAAACCATCAAATTTAGAAGAAAAAGCCCTTTGACTTTTTCCCCCTCTCCCTCCCCAATGGCTGTGTAGCAAACAT
CCCTGGCGATACCTTGGAAAGGACGAAGTTGGTCTGCAGTCGCAATTTCGTGGGTTGAGTTCACAGTTGTGAGTGCG
GGGCTCGGAGATGGAGCCGTGGTCCTCTAGGTGGAAAACGAAACGGTGGCTCTGGGATTTCACCGTAACAACCCTCG
CATTGACCTTCCTCTTCCAAGCTAGAGAGGTCAGAGGAGCTGCTCCAGTTGATGTACTAAAAGCACTAGATTTTCAC
AATTCTCCAGAGGGAATATCAAAAACAACGGGATTTTGCACAAACAGAAAGAATTCTAAAGGCTCAGATACTGCTTA
CAGAGTTTCAAAGCAAGCACAACTCAGTGCCCCAACAAAACAGTTATTTCCAGGTGGAACTTTCCCAGAAGACTTTT
CAATACTATTTACAGTAAAACCAAAAAAAGGAATTCAGTCTTTCCTTTTATCTATATATAATGAGCATGGTATTCAG
CAAATTGGTGTTGAGGTTGGGAGATCACCTGTTTTTCTGTTTGAAGACCACACTGGAAAACCTGCCCCAGAAGACTA
TCCCCTCTTCAGAACTGTTAACATCGCTGACGGGAAGTGGCATCGGGTAGCAATCAGCGTGGAGAAGAAAACTGTGA
CAATGATTGTTGATTGTAAGAAGAAAACCACGAAACCACTTGATAGAAGTGAGAGAGCAATTGTTGATACCAATGGA
ATCACGGTTTTTGGAACAAGGATTTTGGATGAAGAAGTTTTTGAGGGGGACATTCAGCAGTTTTTGATCACAGGTGA
TCCCAAGGCAGCATATGACTACTGTGAGCATTATAGTCCAGACTGTGACTCTTCAGCACCCAAGGCTGCTCAAGCTC
AGGAACCTCAGATAGATGAGGTGAGGAGCACAAGACCAGAGAAGGTCTTCGTATTTCAGTGATATGGACATAGCAGT
GCAGTTTTTGAACTTATACTTATTTATTCCATTTTAATTAAGCTATGCTTTGTATTTTAATTGTGTTGTAATATTTC
CAGGAAAAAGTGACTTGAATATATTTGGTACTTGTTTTC

<210> SEQ ID NO 103

<211> Length : 1,225

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 103

>T11628_PEA_1_T3

GCAAGCCAAAACCCTGGGCAGACTCAATCCAAAAATAAACAATCAAAGAGCATGTTGGCCTGGTCCTTTGCTAGGTA
CTGTAGAGCAGGTGAGAGAGTGAGGGGGAAGGACTCCAAATTAGACCAGTTCTTAGCCATGAAGCAGAGACTCTGAA
GCCAGACTACCTGGGTCCCAATCTTGGGCTTGGTATTTCCTCGCTGTGTGACTCTGGACTGCGCCATGGGGCTCAGC
GACGGGGAATGGCAGTTGGTGCTGAACGTCTGGGGGGAAGGTGGAGGCTGACATCCCAGGCCATGGGCAGGAAGTCCT
CATCAGGCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACCTGAAGTCAGAGGACGAGA
TGAAGGCGTCTGAGGACTTAAAGAAGCATGGTGCCACCGTGCTCACCGCCCTGGGTGGCATCCTTAAGAAGAAGGGG
CATCATGAGGCAGAGATTAAGCCCCTGGCACAGTCGCATGCCACCAAGCACAAGATCCCCGTGAAGTACCTGGAGTT
CATCTCGGAATGCATCATCCAGGTTCTGCAGAGCAAGCATCCCGGGGACTTTGGTGCTGATGCCCAGGGGGCCATGA
ACAAGGCCCTGGAGCTGTTCCGGAAGGACATGGCCTCCAACTACAAGGAGCTGGGCTTCCAGGGCTAGGCCCCTGCC
GCTCCCACCCCCACCCATCTGGGCCCCGGGTTCAAGAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGC
TTCTGAGTGTCTGCTTTGTTTAGTAGAGGTGGGCAGGAGGAGCTGAGGGGCTGGGGCTGGGGTGTTGAAGTTGGCTT
TGCATGCCCAGCGATGCGCCTCCCTGTGGGATGTCATCACCCTGGGAACCGGGAGTGGCCCTTGGCTCACTGTGTTC
TGCATGGTTTGGATCTGAATTAATTGTCCTTTCTTCTAAATCCCAACCGAACTTCTTCCAACCTCCAAACTGGCTGT
AACCCCAAATCCAAGCCATTAACTACACCTGACAGTAGCAATTGTCTGATTAATCACTGGCCCCTTGAAGACAGCAG
AATGTCCCTTTGCAATGAGGAGGAGATCTGGGCTGGGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTG
TGTGCCTCCTCAGGTATGGCAGTGACTCACCTGGTTTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 104

<211> Length : 1,210

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 104

>T11628_PEA_1_T4

TGTCTCTGAGAGCATTGATGAGGTCAGGAAGCCTCCTGTTGGGTAGAGGAGCAACTAAGAGACTGAACTTGGCCCCC
ACCCTGAGGCTCACAAGCTTGAATTGCACCTGAGTTCCAAAGGAGAAGTTGACATTCTTCCAGAACATATGCCCAGT
GTCTTCAACTTGAGATGGAGCTGGGATGCCAAGTCTGCAAATACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAG
TTGGTGCTGAACGTCTGGGGGAAGGTGGAGGCTGACATCCCAGGCCATGGGCAGGAAGTCCTCATCAGGCTCTTTAA
GGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACCTGAAGTCAGAGGACGAGATGAAGGCGTCTGAGG
ACTTAAAGAAGCATGGTGCCACCGTGCTCACCGCCCTGGGTGGCATCCTTAAGAAGAAGGGGCATCATGAGGCAGAG
ATTAAGCCCCTGGCACAGTCGCATGCCACCAAGCACAAGATCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCAT
CATCCAGGTTCTGCAGAGCAAGCATCCCGGGGACTTTGGTGCTGATGCCCAGGGGGCCATGAACAAGGCCCTGGAGC
TGTTCCGGAAGGACATGGCCTCCAACTACAAGGAGCTGGGCTTCCAGGGCTAGGCCCCTGCCGCTCCCACCCCCACC
CATCTGGGCCCCGGGTTCAAGAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGTGTCTGCT

TTGTTTAGTAGAGGTGGGCAGGAGGAGCTGAGGGGCTGGGGCTGGGGTGTTGAAGTTGGCTTTGCATGCCCAGCGAT
GCGCCTCCCTGTGGGATGTCATCACCCTGGGAACCGGGAGTGGCCCTTGGCTCACTGTGTTCTGCATGGTTTGGATC
TGAATTAATTGTCCTTTCTTCTAAATCCCAACCGAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAG
CCATTAACTACACCTGACAGTAGCAATTGTCTGATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCCCTTTGCAA
TGAGGAGGAGATCTGGGCTGGGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCTCCTCAGGT
ATGGCAGTGACTCACCTGGTTTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 105

<211> Length : 1,192

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 105

>T11628_PEA_1_T5

CTTGGCTGGAGGCTCTGCGAGGACAGCTGGGGAGAAGGGGAGCTGTGCCCAGATTTACCAAAGGGAATTGTCAGCTG
TCCAAGGGCTAGCAAATTCCTAGGTCACCTAGATTGGATTTTCTGACCATAAAAACTGTGGGCCAGGTGCACAGCTG
CCTGAGGGGCTCAAACCTGTGCAGACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAGTTGGTGCTGAACGTCTGG
GGGAAGGTGGAGGCTGACATCCCAGGCCATGGGCAGGAAGTCCTCATCAGGCTCTTTAAGGGTCACCCAGAGACTCT
GGAGAAGTTTGACAAGTTCAAGCACCTGAAGTCAGAGGACGAGATGAAGGCGTCTGAGGACTTAAAGAAGCATGGTG
CCACCGTGCTCACCGCCCTGGGTGGCATCCTTAAGAAGAAGGGGCATCATGAGGCAGAGATTAAGCCCCTGGCACAG
TCGCATGCCACCAAGCACAAGATCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCATCATCCAGGTTCTGCAGAG
CAAGCATCCCGGGGACTTTGGTGCTGATGCCCAGGGGGCCATGAACAAGGCCCTGGAGCTGTTCCGGAAGGACATGG
CCTCCAACTACAAGGAGCTGGGCTTCCAGGGCTAGGCCCCTGCCGCTCCCACCCCCACCCATCTGGGCCCCGGGTTC
AAGAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGTGTCTGCTTTGTTTAGTAGAGGTGGG
CAGGAGGAGCTGAGGGGCTGGGGCTGGGGTGTTGAAGTTGGCTTTGCATGCCCAGCGATGCGCCTCCCTGTGGGATG
TCATCACCCTGGGAACCGGGAGTGGCCCTTGGCTCACTGTGTTCTGCATGGTTTGGATCTGAATTAATTGTCCTTTC
TTCTAAATCCCAACCGAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAGCCATTAACTACACCTGAC
AGTAGCAATTGTCTGATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCCCTTTGCAATGAGGAGGAGATCTGGGC
TGGGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCTCCTCAGGTATGGCAGTGACTCACCTG
GTTTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 106

<211> Length : 1,177

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 106

>T11628_PEA_1_T7

CTTGGCTGGAGGCTCTGCGAGGACAGCTGGGGAGAAGGGGAGCTGTGATGGAGGCTCGCTCTGTTGCCAGGCTGGAG
TACAGCGATCTCGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCAAGTAGCTG
GGACTACAGACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAGTTGGTGCTGAACGTCTGGGGGAAGGTGGAGGCT
GACATCCCAGGCCATGGGCAGGAAGTCCTCATCAGGCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAA
GTTCAAGCACCTGAAGTCAGAGGACGAGATGAAGGCGTCTGAGGACTTAAAGAAGCATGGTGCCACCGTGCTCACCG
CCCTGGGTGGCATCCTTAAGAAGAAGGGGCATCATGAGGCAGAGATTAAGCCCCTGGCACAGTCGCATGCCACCAAG
CACAAGATCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCATCATCCAGGTTCTGCAGAGCAAGCATCCCGGGGA
CTTTGGTGCTGATGCCCAGGGGGCCATGAACAAGGCCCTGGAGCTGTTCCGGAAGGACATGGCCTCCAACTACAAGG
AGCTGGGCTTCCAGGGCTAGGCCCCTGCCGCTCCCACCCCCACCCATCTGGGCCCCGGGTTCAAGAGAGAGCGGGGT
CTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGTGTCTGCTTTGTTTAGTAGAGGTGGGCAGGAGGAGCTGAGG
GGCTGGGGCTGGGGTGTTGAAGTTGGCTTTGCATGCCCAGCGATGCGCCTCCCTGTGGGATGTCATCACCCTGGGAA
CCGGGAGTGGCCCTTGGCTCACTGTGTTCTGCATGGTTTGGATCTGAATTAATTGTCCTTTCTTCTAAATCCCAACC
GAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAGCCATTAACTACACCTGACAGTAGCAATTGTCTG
ATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCCCTTTGCAATGAGGAGGAGATCTGGGCTGGGCGGGCCAGCTG
GGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCTCCTCAGGTATGGCAGTGACTCACCTGGTTTTAATAAAACAA
CCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 107

<211> Length : 1,051

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 107

>T11628_PEA_1_T9

GCCTTATTTCTCTGCTGGTTGAGCGAAGGGATTGTCTTCCATGGTCTCCGAGATCCCGTCCCACGCTCATGCCCTAG
AATTCTCTGAGTCCTTGATGCACTTTTGCCTTTGGCGAGGAGGCAGGACAGTCAGGCGTGGAGGCTCTTTAAGGGTC
ACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACCTGAAGTCAGAGGACGAGATGAAGGCGTCTGAGGACTTA
AAGAAGCATGGTGCCACCGTGCTCACCGCCCTGGGTGGCATCCTTAAGAAGAAGGGGCATCATGAGGCAGAGATTAA
GCCCCTGGCACAGTCGCATGCCACCAAGCACAAGATCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCATCATCC
AGGTTCTGCAGAGCAAGCATCCCGGGGACTTTGGTGCTGATGCCCAGGGGGCCATGAACAAGGCCCTGGAGCTGTTC
CGGAAGGACATGGCCTCCAACTACAAGGAGCTGGGCTTCCAGGGCTAGGCCCCTGCCGCTCCCACCCCCACCCATCT

GGGCCCCGGGTTCAAGAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGTGTCTGCTTTGTT
TAGTAGAGGTGGGCAGGAGGAGCTGAGGGGCTGGGGCTGGGGTGTTGAAGTTGGCTTTGCATGCCCAGCGATGCGCC
TCCCTGTGGGATGTCATCACCCTGGGAACCGGGAGTGGCCCTTGGCTCACTGTGTTCTGCATGGTTTGGATCTGAAT
TAATTGTCCTTTCTTCTAAATCCCAACCGAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAGCCATT
AACTACACCTGACAGTAGCAATTGTCTGATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCCCTTTGCAATGAGG
AGGAGATCTGGGCTGGGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCTCCTCAGGTATGGC
AGTGACTCACCTGGTTTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 108

<211> Length : 1,267

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 108

>T11628_PEA_1_T11

TCCTCCCCTTCTTTCCACACGCACAACCACCCCACCCCCTGTGGCCTGAGCTGTCCTGCCTCGCCACAATGGCACCT
GCCCTAAAATAGCTTCCCATGTGAGGGCTAGAGAAAGGAAAAGATTAGACCCTCCCTGGATGAGAGAGAGAAAGTGA
AGGAGGGCAGGGGAGGGGGACAGCGAGCCATTGAGCGATCTTTGTCAAGCATCCCAGAAGGTATAAAAACGCCCTTG
GGACCAGGCAGCCTCAAACCCCAGCTGTTGGGGCCAGGACACCCAGTGAGCCCATACTTGCTCTTTTTGTCTTCTTC
AGACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAGTTGGTGCTGAACGTCTGGGGGAAGGTGGAGGCTGACATCC
CAGGCCATGGGCAGGAAGTCCTCATCAGGCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAG
CACCTGAAGTCAGAGGACGAGATGAAGGCGTCTGAGGACTTAAAGAAGCATGGTGCCACCGTGCTCACCGCCCTGGG
TGGCATCCTTAAGAAGAAGGGGCATCATGAGGCAGAGATTAAGCCCCTGGCACAGTCGCATGCCACCAAGCACAAGA
TCCCCGTGAAGTACCTGGAGTTCATCTCGGAATGCATCATCCAGGTTCTGCAGAGCAAGCATCCCGGGGACTTTGGT
GCTGATGCCCAGGGGGCCATGAACAAGGGCTAGGCCCCTGCCGCTCCCACCCCCACCCATCTGGGCCCCGGGTTCAA
GAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGTGTCTGCTTTGTTTAGTAGAGGTGGGCA
GGAGGAGCTGAGGGGCTGGGGCTGGGGTGTTGAAGTTGGCTTTGCATGCCCAGCGATGCGCCTCCCTGTGGGATGTC
ATCACCCTGGGAACCGGGAGTGGCCCTTGGCTCACTGTGTTCTGCATGGTTTGGATCTGAATTAATTGTCCTTTCTT
CTAAATCCCAACCGAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAGCCATTAACTACACCTGACAG
TAGCAATTGTCTGATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCCCTTTGCAATGAGGAGGAGATCTGGGCTG
GGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCTCCTCAGGTATGGCAGTGACTCACCTGGT
TTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 109

<211> Length : 3,897

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 109

>HUMCEA_PEA_1_T8

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAA
TCCACGCCGTTCAATGTCGCAGAGGGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTA
CAGCTGGTACAAAGGTGAAAGAGTGGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCC
CAGGGCCCGCATACAGTGGTCGAGAGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAGAATGAC
ACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGTATACCC
GGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTG
AACCTGAGACTCAGGACGCAACCTACCTGTGGTGGGTAAACAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTG
TCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACACAGCAAGCTACAAATGTGAAACCCAGAA
CCCAGTGAGTGCCAGGCGCAGTGATTCAGTCATCCTGAATGTCCTCTGTGAGTATATCTGCTCCTCTCTGGCCCAGG
CTGCCAGCCCAAATCCACAGGGCCAGAGGCAGGATTTCTCAGTCCCTCTCAGGTTCAAGTACACAGACCCTCAACCC
TGGACATCCAGACTGTCTGTGACTTTCTGCCCCAGAAAAACCTGGGCAGACCAAGTCTTGACCAAGAATAGGAGGGG
AGGGGCTGCTTCTGTCCTGGGAGGCTCAGGGTCCACACCCTATGATGGGAGAAACAGGTGAATATCTCAGACTCAGG
CTCAGTAGATACAAGAGGGGTTTGGCTGAGACTTTAGGATTGTGATTCAGCTTAGAGGGACACTGTGGTCCTTCCAT
AGACCAGGAACTTCCACTTCCCTCTGACAATATCACCTGTGGCTTTATTTTGTTTGCTCCAGATGGCCCGGATGCCC
CCACCATTTCCCCTCTAAACACATCTTACAGATCAGGGGAAAATCTGAACCTCTCCTGCCACGCAGCCTCTAACCCA
CCTGCACAGTACTCTTGGTTTGTCAATGGGACTTTCCAGCAATCCACCCAAGAGCTCTTTATCCCCAACATCACTGT
GAATAATAGTGGATCCTATACGTGCCAAGCCCATAACTCAGACACTGGCCTCAATAGGACCACAGTCACGACGATCA
CAGTCTATGCAGAGCCACCCAAACCCTTCATCACCAGCAACAACTCCAACCCCGTGGAGGATGAGGATGCTGTAGCC
TTAACCTGTGAACCTGAGATTCAGAACACAACCTACCTGTGGTGGGTAAATAATCAGAGCCTCCCGGTCAGTCCCAG
GCTGCAGCTGTCCAATGACAACAGGACCCTCACTCTACTCAGTGTCACAAGGAATGATGTAGGACCCTATGAGTGTG
GAATCCAGAACGAATTAAGTGTTGACCACAGCGACCCAGTCATCCTGAATGTCCTCTATGGCCCAGACGACCCCACC
ATTTCCCCCTCATACACCTATTACCGTCCAGGGGTGAACCTCAGCCTCTCCTGCCATGCAGCCTCTAACCCACCTGC
ACAGTATTCTTGGCTGATTGATGGGAACATCCAGCAACACACACAAGAGCTCTTTATCTCCAACATCACTGAGAAGA
ACAGCGGACTCTATACCTGCCAGGCCAATAACTCAGCCAGTGGCCACAGCAGGACTACAGTCAAGACAATCACAGTC
TCTGCGGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCAC
CTGTGAACCTGAGGCTCAGAACACAACCTACCTGTGGTGGGTAAATGGTCAGAGCCTCCCAGTCAGTCCCAGGCTGC
AGCTGTCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACGCAAGAGCCTATGTATGTGGAATC
CAGAACTCAGTGAGTGCAAACCGCAGTGACCCAGTCACCCTGGATGTCCTCTATGGGCCGGACACCCCCATCATTTC
CCCCCCAGACTCGTCTTACCTTTCGGGAGCGAACCTCAACCTCTCCTGCCACTCGGCCTCTAACCCATCCCCGCAGT
ATTCTTGGCGTATCAATGGGATACCGCAGCAACACACACAAGTTCTCTTTATCGCCAAAATCACGCCAAATAATAAC

GGGACCTATGCCTGTTTTGTCTCTAACTTGGCTACTGGCCGCAATAATTCCATAGTCAAGAGCATCACAGTCTCTGC
ATCTGGAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGA
TATAGCAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTGACAGTTGTTTTGCTTCTTCCTTAAAGCATTTGCAAC
AGCTACAGTCTAAAATTGCTTCTTTACCAAGGATATTTACAGAAAAGACTCTGACCAGAGATCGAGACCATCCTAGC
CAACATCGTGAAACCCCATCTCTACTAAAAATACAAAAATGAGCTGGGCTTGGTGGCGCGCACCTGTAGTCCCAGTT
ACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGGAGATTGCAGTGAGCCCAGATCGCACCACTGCA
CTCCAGTCTGGCAACAGAGCAAGACTCCATCTCAAAAAGAAAAGAAAAGAAGACTCTGACCTGTACTCTTGAATACA
AGTTTCTGATACCACTGCACTGTCTGAGAATTTCCAAAACTTTAATGAACTAACTGACAGCTTCATGAAACTGTCCA
CCAAGATCAAGCAGAGAAAATAATTAATTTCATGGGACTAAATGAACTAATGAGGATAATATTTTCATAATTTTTTA
TTTGAAATTTTGCTGATTCTTTAAATGTCTTGTTTCCCAGATTTCAGGAAACTTTTTTTCTTTTAAGCTATCCACAG
CTTACAGCAATTTGATAAAATATACTTTTGTGAACAAAAATTGAGACATTTACATTTTCTCCCTATGTGGTCGCTCC
AGACTTGGGAAACTATTCATGAATATTTATATTGTATGGTAATATAGTTATTGCACAAGTTCAATAAAAATCTGCTC
TTTGTATAACAGAATACATTTGAAAACATTGGTTATATTACCAAGACTTTGACTAGAATGTCGTATTTGAGGATATA
AACCCATAGGTAATAAACCCACAGGTACTACAAACAAAGTCTGAAGTCAGCCTTGGTTTGGCTTCCTAGTGTCAATT
AAACTTCTAAAAGTTTAATCTGAGATTCCTTATAAAAACTTCCAGCAAAGCAACTTTAAAAAAGTCTGTGTGGGCCG
GGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGATCCGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCCAGA
CCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAGTTAGCCGGGCGTGGTGGTGGGGGCC
TGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAACGGCATGAACCCGGGAGGCAGGGCTTGCAGTGAGCCAAGA
TCATGCCGCTGCACTCCAGCCTGGGAGACAAAGTGAGACTCCGTCAA

<210> SEQ ID NO 110

<211> Length : 3,347

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 110

>HUMCEA_PEA_1_T9

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAA
TCCACGCCGTTCAATGTCGCAGAGGGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTA
CAGCTGGTACAAAGGTGAAAGAGTGGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCC
CAGGGCCCGCATACAGTGGTCGAGAGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAGAATGAC
ACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGTATACCC
GGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTG
AACCTGAGACTCAGGACGCAACCTACCTGTGGTGGGTAAACAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTG

TCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACACAGCAAGCTACAAATGTGAAACCCAGAA
CCCAGTGAGTGCCAGGCGCAGTGATTCAGTCATCCTGAATGTCCTCTATGGCCCGGATGCCCCCACCATTTCCCCTC
TAAACACATCTTACAGATCAGGGGAAAATCTGAACCTCTCCTGCCACGCAGCCTCTAACCCACCTGCACAGTACTCT
TGGTTTGTCAATGGGACTTTCCAGCAATCCACCCAAGAGCTCTTTATCCCCAACATCACTGTGAATAATAGTGGATC
CTATACGTGCCAAGCCCATAACTCAGACACTGGCCTCAATAGGACCACAGTCACGACGATCACAGTCTATGCAGAGC
CACCCAAACCCTTCATCACCAGCAACAACTCCAACCCCGTGGAGGATGAGGATGCTGTAGCCTTAACCTGTGAACCT
GAGATTCAGAACACAACCTACCTGTGGTGGGTAAATAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTGTCCAA
TGACAACAGGACCCTCACTCTACTCAGTGTCACAAGGAATGATGTAGGACCCTATGAGTGTGGAATCCAGAACGAAT
TAAGTGTTGACCACAGCGACCCAGTCATCCTGAATGTCCTCTATGGCCCAGACGACCCCACCATTTCCCCCTCATAC
ACCTATTACCGTCCAGGGGTGAACCTCAGCCTCTCCTGCCATGCAGCCTCTAACCCACCTGCACAGTATTCTTGGCT
GATTGATGGGAACATCCAGCAACACACACAAGAGCTCTTTATCTCCAACATCACTGAGAAGAACAGCGGACTCTATA
CCTGCCAGGCCAATAACTCAGCCAGTGGCCACAGCAGGACTACAGTCAAGACAATCACAGTCTCTGCGGAGCTGCCC
AAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTGAACCTGAGGC
TCAGAACACAACCTACCTGTGGTGGGTAAATGGTCAGAGCCTCCCAGTCAGTCCCAGGCTGCAGCTGTCCAATGGCA
ACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACGCAAGAGCCTATGTATGTGGAATCCAGAACTCAGTGAGT
GCAAACCGCAGTGACCCAGTCACCCTGGATGTCCTCTATGGGCCGGACACCCCCATCATTTCCCCCCCAGACTCGTC
TTACCTTTCGGGAGCGAACCTCAACCTCTCCTGCCACTCGGCCTCTAACCCATCCCCGCAGTATTCTTGGCGTATCA
ATGGGATACCGCAGCAACACACACAAGTTCTCTTTATCGCCAAAATCACGCCAAATAATAACGGGACCTATGCCTGT
TTTGTCTCTAACTTGGCTACTGGCCGCAATAATTCCATAGTCAAGAGCATCACAGTCTCTGGTAAGTGGCTCCCTGG
AGCATCAGCATCATATTCTGGGGTGGAGTCTATCTGGTTCTCACCAAAGAGCCAAGAAGACATTTTCTTTCCCAGTC
TGTGTTCCATGGGCACAAGGAAATCCCAAATTCTATCCTGAGCCCCCTCACTCCATCTCGGCCAACTCTCTCCTCCC
CGGCTTCTCTGATATCTCACGGCTGACCTCGGGTCCAGCCTGGAATGTGGGGAGGGGCCTCCCTTAGCCCCAGAAGG
CCCCCAATAGTGAAAGGGACTTCATAGTCCAGAAGAAAGAAGGGTCCTTAAGGTCGAGTTGCTCCTCTCTATCACCA
ATATGTCCCTTTCTGTCACCTCTTTGTGTTTTTTCACCTACTCTGTGAGCTACAAGGAACAAGGAGGCTTTGAAACC
AGCCCACACTTTTTCCCCAAATGAGAGGAGGAAGCCCCTTGGATGAGGCAGGAGCAGCTCAGACTCTGCTCCCTGCT
CTGCGCCCGGCTCACCCGGTGACTGGCTCTGCCCTGGCTCCACTTGGGGTGGGACCGGGGCATGTGGAGAAGGTGTC
CAGGTGGCCTGTTTTGAATCTGGGTAAATCAAGCTGCCAATCCACAGCAGAGCCTCCCTTGGGTCAGGTTGCAGGGA
AATGGGAAAAGAGGGAGCCTCGGGACAGACTCCTGAGCTGTGTCCTGGCTCTGAAGTCACTGGCTGTATGAGGCTGT
GGACACAGCACATAGGACACAGCAGAGGAAAGTGAGTGACACACACTTGGAGAAATAGGGAGATTCAGCCATAGGGG
CTCTGCATGGGAGGGAACAGGCAGTGCCAAAAAGTGTGTGTTTATAGAGAGGGTAAGACTATCAGCCACTATATATA
TCTAACATAAAACTTACCATTAACCATTTCTAAGTGTACAATTAAGTGAAACAGCATAAATATCAATCAAGTATATT
GCCCGGTGTGGTGGCTCATCCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGAGTGGATCACCTGAGGTCAGGAGT
TCAAGATACAGAAAAAAAAAAATAGCTAGGCATGGTGGTGGGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCA
GGAGAATCGCTCGAACCTGGGCGGTGTAGTTTGCAGTGAGCCGAGATTGAGCCACTGCACTCCAGCCTGGGTGACAG
AGTGAGACTACATCACAAAAAAAAAAAAAAAAAAGG

<210> SEQ ID NO 111

<211> Length : 1,886

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 111

>HUMCEA_PEA_1_T20

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAA
TCCACGCCGTTCAATGTCGCAGAGGGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTA
CAGCTGGTACAAAGGTGAAAGAGTGGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCC
CAGGGCCCGCATACAGTGGTCGAGAGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAGAATGAC
ACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGTATACCC
GGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTG
AACCTGAGACTCAGGACGCAACCTACCTGTGGTGGGTAAACAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTG
TCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACACAGCAAGCTACAAATGTGAAACCCAGAA
CCCAGTGAGTGCCAGGCGCAGTGATTCAGTCATCCTGAATGTCCTCTATGGCCCGGATGCCCCCACCATTTCCCCTC
TAAACACATCTTACAGATCAGGGGAAAATCTGAACCTCTCCTGCCACGCAGCCTCTAACCCACCTGCACAGTACTCT
TGGTTTGTCAATGGGACTTTCCAGCAATCCACCCAAGAGCTCTTTATCCCCAACATCACTGTGAATAATAGTGGATC
CTATACGTGCCAAGCCCATAACTCAGACACTGGCCTCAATAGGACCACAGTCACGACGATCACAGTCTATGCAGAGC
CACCCAAACCCTTCATCACCAGCAACAACTCCAACCCCGTGGAGGATGAGGATGCTGTAGCCTTAACCTGTGAACCT
GAGATTCAGAACACAACCTACCTGTGGTGGGTAAATAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTGTCCAA
TGACAACAGGACCCTCACTCTACTCAGTGTCACAAGGAATGATGTAGGACCCTATGAGTGTGGAATCCAGAACGAAT
TAAGTGTTGACCACAGCGACCCAGTCATCCTGAATGTCCTCTATGGCCCAGACGACCCCACCATTTCCCCCTCATAC
ACCTATTACCGTCCAGGGGTGAACCTCAGCCTCTCCTGCCATGCAGCCTCTAACCCACCTGCACAGTATTCTTGGCT
GATTGATGGGAACATCCAGCAACACACACAAGAGCTCTTTATCTCCAACATCACTGAGAAGAACAGCGGACTCTATA
CCTGCCAGGCCAATAACTCAGCCAGTGGCCACAGCAGGACTACAGTCAAGACAATCACAGTCTCTGCGGAGCTGCCC
AAGCCCTCCATCTCCAGCAACAACTCCAACCCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTGAACCTGAGGT
TCAGAACACAACCTACCTGTGGTGGGTAAATGGTCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTGTCCAATGGCA
ACATGACCCTCACTCTACTCAGCTGTCAAAAGGAACGATGCAGGATCCTATGAATGTGAAATACAGAACCCAGCGAG
TGCCAACCGCAGTGACCCAGTCACCCTGAATGTCCTCT

<210> SEQ ID NO 112

<211> Length : 2,429

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 112

>HUMCEA_PEA_1_T25

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAA
TCCACGCCGTTCAATGTCGCAGAGGGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTA
CAGCTGGTACAAAGGTGAAAGAGTGGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCC
CAGGGCCCGCATACAGTGGTCGAGAGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAGAATGAC
ACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGTATACCC
GGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTG
AACCTGAGACTCAGGACGCAACCTACCTGTGGTGGGTAAACAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCTG
TCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACACAGCAAGCTACAAATGTGAAACCCAGAA
CCCAGTGAGTGCCAGGCGCAGTGATTCAGTCATCCTGAATGTCCTCTATGGGCCGGACACCCCCATCATTTCCCCCC
CAGACTCGTCTTACCTTTCGGGAGCGAACCTCAACCTCTCCTGCCACTCGGCCTCTAACCCATCCCCGCAGTATTCT
TGGCGTATCAATGGGATACCGCAGCAACACACACAAGTTCTCTTTATCGCCAAAATCACGCCAAATAATAACGGGAC
CTATGCCTGTTTTGTCTCTAACTTGGCTACTGGCCGCAATAATTCCATAGTCAAGAGCATCACAGTCTCTGCATCTG
GAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATATAG
CAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTGACAGTTGTTTTGCTTCTTCCTTAAAGCATTTGCAACAGCTA
CAGTCTAAAATTGCTTCTTTACCAAGGATATTTACAGAAAAGACTCTGACCAGAGATCGAGACCATCCTAGCCAACA
TCGTGAAACCCCATCTCTACTAAAAATACAAAAATGAGCTGGGCTTGGTGGCGCGCACCTGTAGTCCCAGTTACTCG
GGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGGAGATTGCAGTGAGCCCAGATCGCACCACTGCACTCCA
GTCTGGCAACAGAGCAAGACTCCATCTCAAAAAGAAAAGAAAAGAAGACTCTGACCTGTACTCTTGAATACAAGTTT
CTGATACCACTGCACTGTCTGAGAATTTCCAAAACTTTAATGAACTAACTGACAGCTTCATGAAACTGTCCACCAAG
ATCAAGCAGAGAAAATAATTAATTTCATGGGACTAAATGAACTAATGAGGATAATATTTTCATAATTTTTTATTTGA
AATTTTGCTGATTCTTTAAATGTCTTGTTTCCCAGATTTCAGGAAACTTTTTTTCTTTTAAGCTATCCACAGCTTAC
AGCAATTTGATAAAATATACTTTTGTGAACAAAAATTGAGACATTTACATTTTCTCCCTATGTGGTCGCTCCAGACT
TGGGAAACTATTCATGAATATTTATATTGTATGGTAATATAGTTATTGCACAAGTTCAATAAAAATCTGCTCTTTGT
ATAACAGAATACATTTGAAAACATTGGTTATATTACCAAGACTTTGACTAGAATGTCGTATTTGAGGATATAAACCC
ATAGGTAATAAACCCACAGGTACTACAAACAAAGTCTGAAGTCAGCCTTGGTTTGGCTTCCTAGTGTCAATTAAACT
TCTAAAAGTTTAATCTGAGATTCCTTATAAAAACTTCCAGCAAAGCAACTTTAAAAAAGTCTGTGTGGGCCGGGCGC
GGTGGCTCACGCCTGTAATCCCAGCACTTTGATCCGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCCAGACCATC
CTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAGTTAGCCGGGCGTGGTGGTGGGGGCCTGTAG
TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAACGGCATGAACCCGGGAGGCAGGGCTTGCAGTGAGCCAAGATCATG
CCGCTGCACTCCAGCCTGGGAGACAAAGTGAGACTCCGTCAA

<210> SEQ ID NO 113

<211> Length : 2,429

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 113

>HUMCEA_PEA_1_T26

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAGCCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAA
TCCACGCCGTTCAATGTCGCAGAGGGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTA
CAGCTGGTACAAAGGTGAAAGAGTGGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCC
CAGGGCCCGCATACAGTGGTCGAGAGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAGAATGAC
ACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGTATACCC
GGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTG
AACCTGAGGCTCAGAACACAACCTACCTGTGGTGGGTAAATGGTCAGAGCCTCCCAGTCAGTCCCAGGCTGCAGCTG
TCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACGCAAGAGCCTATGTATGTGGAATCCAGAA
CTCAGTGAGTGCAAACCGCAGTGACCCAGTCACCCTGGATGTCCTCTATGGGCCGGACACCCCCATCATTTCCCCCC
CAGACTCGTCTTACCTTTCGGGAGCGAACCTCAACCTCTCCTGCCACTCGGCCTCTAACCCATCCCCGCAGTATTCT
TGGCGTATCAATGGGATACCGCAGCAACACACACAAGTTCTCTTTATCGCCAAAATCACGCCAAATAATAACGGGAC
CTATGCCTGTTTTGTCTCTAACTTGGCTACTGGCCGCAATAATTCCATAGTCAAGAGCATCACAGTCTCTGCATCTG
GAACTTCTCCTGGTCTCTCAGCTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATATAG
CAGCCCTGGTGTAGTTTCTTCATTTCAGGAAGACTGACAGTTGTTTTGCTTCTTCCTTAAAGCATTTGCAACAGCTA
CAGTCTAAAATTGCTTCTTTACCAAGGATATTTACAGAAAAGACTCTGACCAGAGATCGAGACCATCCTAGCCAACA
TCGTGAAACCCCATCTCTACTAAAAATACAAAAATGAGCTGGGCTTGGTGGCGCGCACCTGTAGTCCCAGTTACTCG
GGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGGAGATTGCAGTGAGCCCAGATCGCACCACTGCACTCCA
GTCTGGCAACAGAGCAAGACTCCATCTCAAAAAGAAAAGAAAAGAAGACTCTGACCTGTACTCTTGAATACAAGTTT
CTGATACCACTGCACTGTCTGAGAATTTCCAAAACTTTAATGAACTAACTGACAGCTTCATGAAACTGTCCACCAAG
ATCAAGCAGAGAAAATAATTAATTTCATGGGACTAAATGAACTAATGAGGATAATATTTTCATAATTTTTTATTTGA
AATTTTGCTGATTCTTTAAATGTCTTGTTTCCCAGATTTCAGGAAACTTTTTTTCTTTTAAGCTATCCACAGCTTAC
AGCAATTTGATAAAATATACTTTTGTGAACAAAAATTGAGACATTTACATTTTCTCCCTATGTGGTCGCTCCAGACT
TGGGAAACTATTCATGAATATTTATATTGTATGGTAATATAGTTATTGCACAAGTTCAATAAAAATCTGCTCTTTGT
ATAACAGAATACATTTGAAAACATTGGTTATATTACCAAGACTTTGACTAGAATGTCGTATTTGAGGATATAAACCC
ATAGGTAATAAACCCACAGGTACTACAAACAAAGTCTGAAGTCAGCCTTGGTTTGGCTTCCTAGTGTCAATTAAACT
TCTAAAAGTTTAATCTGAGATTCCTTATAAAAACTTCCAGCAAAGCAACTTTAAAAAAGTCTGTGTGGGCCGGGCGC
GGTGGCTCACGCCTGTAATCCCAGCACTTTGATCCGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCCAGACCATC

CTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAGTTAGCCGGGCGTGGTGGTGGGGGCCTGTAG
TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAACGGCATGAACCCGGGAGGCAGGGCTTGCAGTGAGCCAAGATCATG
CCGCTGCACTCCAGCCTGGGAGACAAAGTGAGACTCCGTCAA

<210> SEQ ID NO 114

<211> Length : 3,898

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 114

>R35137_PEA_1_PEA_1_PEA_1_T3

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG
GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAA
CCCCACCGGGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGC
TGGCGGACGAGGTGCGCGGCGCGGGGGAGCGGGAAGCCGGGCAACAGTCCGCCCCCGTGACGCCTTGCGCCCTTCCA
GGTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAGATGGGGCCGC
CCTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGAGTGCGGGTTCCGCGGC
GGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAAGCTGATGAGTGTGCGGCTGTGCCC
GCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGCCCGCGCCCACCGACCCCTCCTTTGCGCAGTTCC
AGGCTGAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCCAAGCTCACCGAGCAGGTCTTCAATGAGGCTCCT
GGCATCAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCCGCGCGTGCAGCTGCCCCCGCGGGCGGTGGAGCG
CGCTCAGGAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCCTCCTGGAGGAGACCGGCATCTGCGTGGTGC
CAGGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGATGACCATTCTGCCCCCCTTGGAGAAACTGCGG
CTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGTACTCCTGAGCACCCCAGCTGGGGCCAGG

CTGGGTCGCCCTGGACTGTGTGCTCAGGAGCCCTGGGAGGCTCTGGAGCCCACTGTACTTGCTCTTGATGCCTGGCG
GGGTGGGGTGGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCTAATAAAGCTGTGTGGCAGTCTGACTCCAAA
AAGGAAGCGTTGGCAGCTGCGTGGCCCGCTCCCACCTGCCTACCCTTCTTGCAGGCCTGAGTCCCTTCAGAGAAGGG
ACCTTCCACGGCCACCACCCACCTCTTCCTCCTGAAGACCCCGTGCCCACCATAGGCTGGGTCTTCCCTCTGGCCTC
TGGTTGTGGGGCAGAGCCCGTCAGATCACACAGAAATGGGTTGAGAGGGTCCAGAGTGTGAGGAAAGCGCAGGCCCC
ACACCCCTTTGTGGAAGCCCCCAAGAATCTAGGGAGCCAGGGGCCCAGGTGGCCACCCGAAGAAACACAGCCTTTCC
TGAGGAAGGCACAGTGAACTGCCTCCTTCCTGGCTCCCTTTCCTGTGAGGTCCATGTCTTCCCTGGGGCAGGGGGAA
ATACTAAACCAGCATGGTGCTGGCTGGTCAGGGTGACTGACAGCTCAGGAAGGAAGTCTTGGTTCTCTTACCCAAGG
AAGCAGGGGTGGGGCCACTGTCTGGGGGGCCAGAGACCACCTTTGGTGTCATTGTGTGGTGCAGTCCTTCGGCTGGG
TGGAGTGGGAGGCAGAGGGAGAGGATAAGGGAGCGTCCCAGGGGAGGGCTGGGGCTGGAGGAGGCAGGGGCTGGGCT
GAGCGGGAGTGGGCAGCGCTGTGTCCTGGCCTGGGGAGCATGGCTGAGCACCTACTACATGCAGACACTGCTGGGGG
GTCACTCAATCTGCACAGATGCTCTTCTGAAGTAGGCATGATAGTCCCCATTTGATAGACGTGGAAACCTGCAACCC
AAAAACCTGCTGAGCTGACAAGAAACCCCCTCAGAGGCCTCGGCAGCCAGAAAAATGCGTTGGGTCCAGTGCCCTCA
AGTCCGCCAAGGACATGGGCTGGCTTTAGAGACTCACAAACTTGGGAGATAGGACTGGCCAAGGGCACCTGGTTTTT
TTCGCTCTGGAGATGGTTCTTAACCACAGGCCACACACACTTCACAGCCTCATCTGGCCCTCGGGAGCCCCAGAGGG
CACAGCTCTGGGCAGGAGACACAGCAGGTGGGCCCCTCCCTTGGCAGGGCGGGCTCGAATCAGGCAGGGTGCTCCTA
GCCTTGTCACCGGACACCGAAGGGCGTCACGGGCAGTGGCTGGCGGTGTCCTTCTGAGCTAAGCTGGGTCCTGACCT
TTCACACTTCCCTCCTAACTTCCATGGCTCTGTCACCGCCTTACGGAGGAGCTGAAGCCACAGACACAGCAAGGTTG
GGGTCCGCACCGGAAGTATCCAGTGGTAGACGGCGGAACCCTTAAGAAACGGACGCCTTCATGCGGGCGGCTGGAGA
AGCGGGGGCTGGGCACTGCAGCAACCACGCTTCGGCTGACACCAGGAAGGAAGCACGCCTGGAGCGGATCCGAAACT
ACTGAGAGGGGCCAGGGCTGGCCGTGGGCGCAGGCCGATCCTTACATTCGAGGCCGGCCCAGCTCTGTAGCTTCCCC
CTCTGGGCCTCTCACCCGCGCAGGACCTCGGTGGGAGCGCGCACGTGGCGGGGCGGGGGGCCGCGGGCCCGAGCCCG
GACTGGCCACCGGGGGCGCCCGCGAGCTGACGCTCTGGCCCGTTGCGGTCTCTGTGGCCCGCGCGACCTTCCGGCCC
TGGAGCCGGTGGCCGCGGGGCTCCAGCGACGCCGTGTGGTCCGTGCTCCGCTCTGTGGCTCCAGGGGGCCGAGAAAC
TGCTGAGAGTCCGGCCCGGCCGGCAGTGCTGGGCGCGGGCCGAGGGCCCCGGGAGGCAGCGGCCCCGCCCTCTTTAC
CTGCGGCCTCGCAGAGCATGCTGGGAGCCGCGGGAGGCAGTGGCCCCGCTCCCCTCACCTGCGGTATCGCAGAGCAT
GGTGGGAGCCCCGGGAGGCAGTGGCCCCGCCCCCTTCCCTGCGGCCGC

<210> SEQ ID NO 115

<211> Length : 3,774

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 115

>R35137_PEA_1_PEA_1_PEA_1_T5

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG
GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAA
CCCCACCGGGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGC
TGGCGGACGAGGTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAG
ATGGGGCCGCCCTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGAGTGCGG
GTTCCGCGGCGGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAAGCTGATGAGTGTGC
GGCTGTGCCCGCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGCCCGCGCCCACCGACCCCTCCTTT
GCGCAGTTCCAGGCTGAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCCAAGCTCACCGAGCAGGTCTTCAA
TGAGGCTCCTGGCATCAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCCGCGCGTGCAGCTGCCCCCGCGGG
CGGTGGAGCGCGCTCAGGAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCCTCCTGGAGGAGACCGGCATC
TGCGTGGTGCCAGGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGATGACCATTCTGCCCCCCTTGGA
GAAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGAGCCCTGGGAGGCTCTGGA
GCCCACTGTACTTGCTCTTGATGCCTGGCGGGGTGGGGTGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCT
AATAAAGCTGTGTGGCAGTCTGACTCCAAAAAGGAAGCGTTGGCAGCTGCGTGGCCCGCTCCCACCTGCCTACCCTT
CTTGCAGGCCTGAGTCCCTTCAGAGAAGGGACCTTCCACGGCCACCACCCACCTCTTCCTCCTGAAGACCCCGTGCC
CACCATAGGCTGGGTCTTCCCTCTGGCCTCTGGTTGTGGGGCAGAGCCCGTCAGATCACACAGAAATGGGTTGAGAG
GGTCCAGAGTGTGAGGAAAGCGCAGGCCCCACACCCCTTTGTGGAAGCCCCCAAGAATCTAGGGAGCCAGGGGCCCA
GGTGGCCACCCGAAGAAACACAGCCTTTCCTGAGGAAGGCACAGTGAACTGCCTCCTTCCTGGCTCCCTTTCCTGTG
AGGTCCATGTCTTCCCTGGGGCAGGGGGAAATACTAAACCAGCATGGTGCTGGCTGGTCAGGGTGACTGACAGCTCA
GGAAGGAAGTCTTGGTTCTCTTACCCAAGGAAGCAGGGGTGGGGCCACTGTCTGGGGGGCCAGAGACCACCTTTGGT
GTCATTGTGTGGTGCAGTCCTTCGGCTGGGTGGAGTGGGAGGCAGAGGGAGAGGATAAGGGAGCGTCCCAGGGGAGG
GCTGGGGCTGGAGGAGGCAGGGGCTGGGCTGAGCGGGAGTGGGCAGCGCTGTGTCCTGGCCTGGGGAGCATGGCTGA
GCACCTACTACATGCAGACACTGCTGGGGGGTCACTCAATCTGCACAGATGCTCTTCTGAAGTAGGCATGATAGTCC
CCATTTGATAGACGTGGAAACCTGCAACCCAAAAACCTGCTGAGCTGACAAGAAACCCCCTCAGAGGCCTCGGCAGC
CAGAAAAATGCGTTGGGTCCAGTGCCCTCAAGTCCGCCAAGGACATGGGCTGGCTTTAGAGACTCACAAACTTGGGA
GATAGGACTGGCCAAGGGCACCTGGTTTTTTTCGCTCTGGAGATGGTTCTTAACCACAGGCCACACACACTTCACAG
CCTCATCTGGCCCTCGGGAGCCCCAGAGGGCACAGCTCTGGGCAGGAGACACAGCAGGTGGGCCCCTCCCTTGGCAG

GGCGGGCTCGAATCAGGCAGGGTGCTCCTAGCCTTGTCACCGGACACCGAAGGGCGTCACGGGCAGTGGCTGGCGGT
GTCCTTCTGAGCTAAGCTGGGTCCTGACCTTTCACACTTCCCTCCTAACTTCCATGGCTCTGTCACCGCCTTACGGA
GGAGCTGAAGCCACAGACACAGCAAGGTTGGGGTCCGCACCGGAAGTATCCAGTGGTAGACGGCGGAACCCTTAAGA
AACGGACGCCTTCATGCGGGCGGCTGGAGAAGCGGGGGCTGGGCACTGCAGCAACCACGCTTCGGCTGACACCAGGA
AGGAAGCACGCCTGGAGCGGATCCGAAACTACTGAGAGGGGCCAGGGCTGGCCGTGGGCGCAGGCCGATCCTTACAT
TCGAGGCCGGCCCAGCTCTGTAGCTTCCCCCTCTGGGCCTCTCACCCGCGCAGGACCTCGGTGGGAGCGCGCACGTG
GCGGGGCGGGGGGCCGCGGGCCCGAGCCCGGACTGGCCACCGGGGGCGCCCGCGAGCTGACGCTCTGGCCCGTTGCG
GTCTCTGTGGCCCGCGCGACCTTCCGGCCCTGGAGCCGGTGGCCGCGGGGCTCCAGCGACGCCGTGTGGTCCGTGCT
CCGCTCTGTGGCTCCAGGGGGCCGAGAAACTGCTGAGAGTCCGGCCCGGCCGGCAGTGCTGGGCGCGGGCCGAGGGC
CCCGGGAGGCAGCGGCCCCGCCCTCTTTACCTGCGGCCTCGCAGAGCATGCTGGGAGCCGCGGGAGGCAGTGGCCCC
GCTCCCCTCACCTGCGGTATCGCAGAGCATGGTGGGAGCCCCGGGAGGCAGTGGCCCCGCCCCCTTCCCTGCGGCCG
C

<210> SEQ ID NO 116

<211> Length : 3,998

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 116

>R35137_PEA_1_PEA_1_PEA_1_T10

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG
GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAA
CCCCACCGGGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGC
TGGCGGACGAGGTGCGCGGCGCGGGGGAGCGGGAAGCCGGGCAACAGTCCGCCCCCGTGACGCCTTGCGCCCTTCCA
GGTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAGATGGGGCCGC

CCTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGAGTGCGTGCGTACGAGG
CGGGTGGGGGCTCGCGGGCCATGGCCAGGCCCTCCTCGCCCGATGGGCCACCCCCTCCTCCGCACCTGACCTGGCCG
TGCGCAGGTGCGGGTTCCGCGGCGGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAAG
CTGATGAGTGTGCGGCTGTGCCCGCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGCCCGCGCCCAC
CGACCCCTCCTTTGCGCAGTTCCAGGCTGAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCCAAGCTCACCG
AGCAGGTCTTCAATGAGGCTCCTGGCATCAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCCGCGCGTGCAG
CTGCCCCCGCGGGCGGTGGAGCGCGCTCAGGAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCCTCCTGGA
GGAGACCGGCATCTGCGTGGTGCCAGGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGATGACCATTC
TGCCCCCCTTGGAGAAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGTACTCC
TGAGCACCCCAGCTGGGGCCAGGCTGGGTCGCCCTGGACTGTGTGCTCAGGAGCCCTGGGAGGCTCTGGAGCCCACT
GTACTTGCTCTTGATGCCTGGCGGGGTGGGGTGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCTAATAAAG
CTGTGTGGCAGTCTGACTCCAAAAAGGAAGCGTTGGCAGCTGCGTGGCCCGCTCCCACCTGCCTACCCTTCTTGCAG
GCCTGAGTCCCTTCAGAGAAGGGACCTTCCACGGCCACCACCCACCTCTTCCTCCTGAAGACCCCGTGCCCACCATA
GGCTGGGTCTTCCCTCTGGCCTCTGGTTGTGGGGCAGAGCCCGTCAGATCACACAGAAATGGGTTGAGAGGGTCCAG
AGTGTGAGGAAAGCGCAGGCCCCACACCCCTTTGTGGAAGCCCCCAAGAATCTAGGGAGCCAGGGGCCCAGGTGGCC
ACCCGAAGAAACACAGCCTTTCCTGAGGAAGGCACAGTGAACTGCCTCCTTCCTGGCTCCCTTTCCTGTGAGGTCCA
TGTCTTCCCTGGGGCAGGGGGAAATACTAAACCAGCATGGTGCTGGCTGGTCAGGGTGACTGACAGCTCAGGAAGGA
AGTCTTGGTTCTCTTACCCAAGGAAGCAGGGGTGGGGCCACTGTCTGGGGGGCCAGAGACCACCTTTGGTGTCATTG
TGTGGTGCAGTCCTTCGGCTGGGTGGAGTGGGAGGCAGAGGGAGAGGATAAGGGAGCGTCCCAGGGGAGGGCTGGGG
CTGGAGGAGGCAGGGGCTGGGCTGAGCGGGAGTGGGCAGCGCTGTGTCCTGGCCTGGGGAGCATGGCTGAGCACCTA
CTACATGCAGACACTGCTGGGGGGTCACTCAATCTGCACAGATGCTCTTCTGAAGTAGGCATGATAGTCCCCATTTG
ATAGACGTGGAAACCTGCAACCCAAAAACCTGCTGAGCTGACAAGAAACCCCCTCAGAGGCCTCGGCAGCCAGAAAA
ATGCGTTGGGTCCAGTGCCCTCAAGTCCGCCAAGGACATGGGCTGGCTTTAGAGACTCACAAACTTGGGAGATAGGA
CTGGCCAAGGGCACCTGGTTTTTTTCGCTCTGGAGATGGTTCTTAACCACAGGCCACACACACTTCACAGCCTCATC
TGGCCCTCGGGAGCCCCAGAGGGCACAGCTCTGGGCAGGAGACACAGCAGGTGGGCCCCTCCCTTGGCAGGGCGGGC
TCGAATCAGGCAGGGTGCTCCTAGCCTTGTCACCGGACACCGAAGGGCGTCACGGGCAGTGGCTGGCGGTGTCCTTC
TGAGCTAAGCTGGGTCCTGACCTTTCACACTTCCCTCCTAACTTCCATGGCTCTGTCACCGCCTTACGGAGGAGCTG
AAGCCACAGACACAGCAAGGTTGGGGTCCGCACCGGAAGTATCCAGTGGTAGACGGCGGAACCCTTAAGAAACGGAC
GCCTTCATGCGGGCGGCTGGAGAAGCGGGGGCTGGGCACTGCAGCAACCACGCTTCGGCTGACACCAGGAAGGAAGC
ACGCCTGGAGCGGATCCGAAACTACTGAGAGGGGCCAGGGCTGGCCGTGGGCGCAGGCCGATCCTTACATTCGAGGC
CGGCCCAGCTCTGTAGCTTCCCCCTCTGGGCCTCTCACCCGCGCAGGACCTCGGTGGGAGCGCGCACGTGGCGGGGC
GGGGGGCCGCGGGCCCGAGCCCGGACTGGCCACCGGGGGCGCCCGCGAGCTGACGCTCTGGCCCGTTGCGGTCTCTG
TGGCCCGCGCGACCTTCCGGCCCTGGAGCCGGTGGCCGCGGGGCTCCAGCGACGCCGTGTGGTCCGTGCTCCGCTCT
GTGGCTCCAGGGGGCCGAGAAACTGCTGAGAGTCCGGCCCGGCCGGCAGTGCTGGGCGCGGGCCGAGGGCCCCGGGA
GGCAGCGGCCCCGCCCTCTTTACCTGCGGCCTCGCAGAGCATGCTGGGAGCCGCGGGAGGCAGTGGCCCCGCTCCCC
TCACCTGCGGTATCGCAGAGCATGGTGGGAGCCCCGGGAGGCAGTGGCCCCGCCCCCTTCCCTGCGGCCGC

<210> SEQ ID NO 117

<211> Length : 4,071

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 117

>R35137_PEA_1_PEA_1_PEA_1_T11

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG
GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAA
CCCCACCGGGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGC
TGGCGGACGAGGTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAG
ATGGGGCCGCCCTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGAGTGCGT
GCGTACGAGGCGGGTGGGGGCTCGCGGGCCATGGCCAGGCCCTCCTCGCCCGATGGGCCACCCCCTCCTCCGCACCT
GACCTGGCCGTGCGCAGGTGCGGGTTCCGCGGCGGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCA
GATGCTGAAGCTGATGAGTGTGCGGCTGTGCCCGCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGC
CCGCGCCCACCGACCCCTCCTTTGCGCAGTTCCAGGCTGAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCC
AAGCTCACCGAGCAGGTCTTCAATGAGGCTCCTGGCATCAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCC
GCGCGTGCAGCTGCCCCCGCGGGCGGTGGAGCGCGCTCAGGTCAGGCGGGGGCGGGGCCTGCGGGGTGGGCAGGGGG
GGCCGGGCATCCCTCTCTGACGGCTCTCCGTCCACAGGAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCC
TCCTGGAGGAGACCGGCATCTGCGTGGTGCCAGGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGTGA
GGCCTGGCCCTCACTCCCTGTCCCGCCACCCTGGCCCTTCACTCACTGTCAACTCCTTTCAGGATGACCATTCTGCC
CCCCTTGGAGAAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGTACTCCTGAG
CACCCCAGCTGGGGCCAGGCTGGGTCGCCCTGGACTGTGTGCTCAGGAGCCCTGGGAGGCTCTGGAGCCCACTGTAC
TTGCTCTTGATGCCTGGCGGGGTGGGGTGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCTAATAAAGCTGT
GTGGCAGTCTGACTCCAAAAAGGAAGCGTTGGCAGCTGCGTGGCCCGCTCCCACCTGCCTACCCTTCTTGCAGGCCT

GAGTCCCTTCAGAGAAGGGACCTTCCACGGCCACCACCCACCTCTTCCTCCTGAAGACCCCGTGCCCACCATAGGCT
GGGTCTTCCCTCTGGCCTCTGGTTGTGGGGCAGAGCCCGTCAGATCACACAGAAATGGGTTGAGAGGGTCCAGAGTG
TGAGGAAAGCGCAGGCCCCACACCCCTTTGTGGAAGCCCCCAAGAATCTAGGGAGCCAGGGGCCCAGGTGGCCACCC
GAAGAAACACAGCCTTTCCTGAGGAAGGCACAGTGAACTGCCTCCTTCCTGGCTCCCTTTCCTGTGAGGTCCATGTC
TTCCCTGGGGCAGGGGGAAATACTAAACCAGCATGGTGCTGGCTGGTCAGGGTGACTGACAGCTCAGGAAGGAAGTC
TTGGTTCTCTTACCCAAGGAAGCAGGGGTGGGGCCACTGTCTGGGGGGCCAGAGACCACCTTTGGTGTCATTGTGTG
GTGCAGTCCTTCGGCTGGGTGGAGTGGGAGGCAGAGGGAGAGGATAAGGGAGCGTCCCAGGGGAGGGCTGGGGCTGG
AGGAGGCAGGGGCTGGGCTGAGCGGGAGTGGGCAGCGCTGTGTCCTGGCCTGGGGAGCATGGCTGAGCACCTACTAC
ATGCAGACACTGCTGGGGGGTCACTCAATCTGCACAGATGCTCTTCTGAAGTAGGCATGATAGTCCCCATTTGATAG
ACGTGGAAACCTGCAACCCAAAAACCTGCTGAGCTGACAAGAAACCCCCTCAGAGGCCTCGGCAGCCAGAAAAATGC
GTTGGGTCCAGTGCCCTCAAGTCCGCCAAGGACATGGGCTGGCTTTAGAGACTCACAAACTTGGGAGATAGGACTGG
CCAAGGGCACCTGGTTTTTTTCGCTCTGGAGATGGTTCTTAACCACAGGCCACACACACTTCACAGCCTCATCTGGC
CCTCGGGAGCCCCAGAGGGCACAGCTCTGGGCAGGAGACACAGCAGGTGGGCCCCTCCCTTGGCAGGGCGGGCTCGA
ATCAGGCAGGGTGCTCCTAGCCTTGTCACCGGACACCGAAGGGCGTCACGGGCAGTGGCTGGCGGTGTCCTTCTGAG
CTAAGCTGGGTCCTGACCTTTCACACTTCCCTCCTAACTTCCATGGCTCTGTCACCGCCTTACGGAGGAGCTGAAGC
CACAGACACAGCAAGGTTGGGGTCCGCACCGGAAGTATCCAGTGGTAGACGGCGGAACCCTTAAGAAACGGACGCCT
TCATGCGGGCGGCTGGAGAAGCGGGGGCTGGGCACTGCAGCAACCACGCTTCGGCTGACACCAGGAAGGAAGCACGC
CTGGAGCGGATCCGAAACTACTGAGAGGGGCCAGGGCTGGCCGTGGGCGCAGGCCGATCCTTACATTCGAGGCCGGC
CCAGCTCTGTAGCTTCCCCCTCTGGGCCTCTCACCCGCGCAGGACCTCGGTGGGAGCGCGCACGTGGCGGGGCGGGG
GGCCGCGGGCCCGAGCCCGGACTGGCCACCGGGGGCGCCCGCGAGCTGACGCTCTGGCCCGTTGCGGTCTCTGTGGC
CCGCGCGACCTTCCGGCCCTGGAGCCGGTGGCCGCGGGGCTCCAGCGACGCCGTGTGGTCCGTGCTCCGCTCTGTGG
CTCCAGGGGGCCGAGAAACTGCTGAGAGTCCGGCCCGGCCGGCAGTGCTGGGCGCGGGCCGAGGGCCCCGGGAGGCA
GCGGCCCCGCCCTCTTTACCTGCGGCCTCGCAGAGCATGCTGGGAGCCGCGGGAGGCAGTGGCCCCGCTCCCCTCAC
CTGCGGTATCGCAGAGCATGGTGGGAGCCCCGGGAGGCAGTGGCCCCGCCCCCTTCCCTGCGGCCGC

<210> SEQ ID NO 118

<211> Length : 4,138

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 118

>R35137_PEA_1_PEA_1_PEA_1_T12

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG

GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAA
CCCCACCGGGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGC
TGGCGGACGAGGTGCGCGGCGCGGGGGAGCGGGAAGCCGGGCAACAGTCCGCCCCCGTGACGCCTTGCGCCCTTCCA
GGTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAGATGGGGCCGC
CCTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGAGTGCGTGCGTACGAGG
CGGGTGGGGGCTCGCGGGCCATGGCCAGGCCCTCCTCGCCCGATGGGCCACCCCCTCCTCCGCACCTGACCTGGCCG
TGCGCAGGTGCGGGTTCCGCGGCGGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAAG
CTGATGAGTGTGCGGCTGTGCCCGCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGCCCGCGCCCAC
CGACCCCTCCTTTGCGCAGTTCCAGGCTGAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCCAAGCTCACCG
AGCAGGTCTTCAATGAGGCTCCTGGCATCAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCCGCGCGTGCAG
CTGCCCCCGCGGGCGGTGGAGCGCGCTCAGGTCAGGCGGGGGCGGGGCCTGCGGGGTGGGCAGGGGGGGCCGGGCAT
CCCTCTCTGACGGCTCTCCGTCCACAGGAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCCTCCTGGAGGA
GACCGGCATCTGCGTGGTGCCAGGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGTGAGGCCTGGCCC
TCACTCCCTGTCCCGCCACCCTGGCCCTTCACTCACTGTCAACTCCTTTCAGGATGACCATTCTGCCCCCCTTGGAG
AAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGTACTCCTGAGCACCCCAGCT
GGGGCCAGGCTGGGTCGCCCTGGACTGTGTGCTCAGGAGCCCTGGGAGGCTCTGGAGCCCACTGTACTTGCTCTTGA
TGCCTGGCGGGGTGGGGTGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCTAATAAAGCTGTGTGGCAGTCT
GACTCCAAAAAGGAAGCGTTGGCAGCTGCGTGGCCCGCTCCCACCTGCCTACCCTTCTTGCAGGCCTGAGTCCCTTC
AGAGAAGGGACCTTCCACGGCCACCACCCACCTCTTCCTCCTGAAGACCCCGTGCCCACCATAGGCTGGGTCTTCCC
TCTGGCCTCTGGTTGTGGGGCAGAGCCCGTCAGATCACACAGAAATGGGTTGAGAGGGTCCAGAGTGTGAGGAAAGC
GCAGGCCCCACACCCCTTTGTGGAAGCCCCCAAGAATCTAGGGAGCCAGGGGCCCAGGTGGCCACCCGAAGAAACAC
AGCCTTTCCTGAGGAAGGCACAGTGAACTGCCTCCTTCCTGGCTCCCTTTCCTGTGAGGTCCATGTCTTCCCTGGGG
CAGGGGGAAATACTAAACCAGCATGGTGCTGGCTGGTCAGGGTGACTGACAGCTCAGGAAGGAAGTCTTGGTTCTCT
TACCCAAGGAAGCAGGGGTGGGGCCACTGTCTGGGGGGCCAGAGACCACCTTTGGTGTCATTGTGTGGTGCAGTCCT
TCGGCTGGGTGGAGTGGGAGGCAGAGGGAGAGGATAAGGGAGCGTCCCAGGGGAGGGCTGGGGCTGGAGGAGGCAGG
GGCTGGGCTGAGCGGGAGTGGGCAGCGCTGTGTCCTGGCCTGGGGAGCATGGCTGAGCACCTACTACATGCAGACAC
TGCTGGGGGGTCACTCAATCTGCACAGATGCTCTTCTGAAGTAGGCATGATAGTCCCCATTTGATAGACGTGGAAAC
CTGCAACCCAAAAACCTGCTGAGCTGACAAGAAACCCCCTCAGAGGCCTCGGCAGCCAGAAAAATGCGTTGGGTCCA
GTGCCCTCAAGTCCGCCAAGGACATGGGCTGGCTTTAGAGACTCACAAACTTGGGAGATAGGACTGGCCAAGGGCAC
CTGGTTTTTTTCGCTCTGGAGATGGTTCTTAACCACAGGCCACACACACTTCACAGCCTCATCTGGCCCTCGGGAGC

CCCAGAGGGCACAGCTCTGGGCAGGAGACACAGCAGGTGGGCCCCTCCCTTGGCAGGGCGGGCTCGAATCAGGCAGG
GTGCTCCTAGCCTTGTCACCGGACACCGAAGGGCGTCACGGGCAGTGGCTGGCGGTGTCCTTCTGAGCTAAGCTGGG
TCCTGACCTTTCACACTTCCCTCCTAACTTCCATGGCTCTGTCACCGCCTTACGGAGGAGCTGAAGCCACAGACACA
GCAAGGTTGGGGTCCGCACCGGAAGTATCCAGTGGTAGACGGCGGAACCCTTAAGAAACGGACGCCTTCATGCGGGC
GGCTGGAGAAGCGGGGGCTGGGCACTGCAGCAACCACGCTTCGGCTGACACCAGGAAGGAAGCACGCCTGGAGCGGA
TCCGAAACTACTGAGAGGGGCCAGGGCTGGCCGTGGGCGCAGGCCGATCCTTACATTCGAGGCCGGCCCAGCTCTGT
AGCTTCCCCCTCTGGGCCTCTCACCCGCGCAGGACCTCGGTGGGAGCGCGCACGTGGCGGGGCGGGGGGCCGCGGGC
CCGAGCCCGGACTGGCCACCGGGGGCGCCCGCGAGCTGACGCTCTGGCCCGTTGCGGTCTCTGTGGCCCGCGCGACC
TTCCGGCCCTGGAGCCGGTGGCCGCGGGGCTCCAGCGACGCCGTGTGGTCCGTGCTCCGCTCTGTGGCTCCAGGGGG
CCGAGAAACTGCTGAGAGTCCGGCCCGGCCGGCAGTGCTGGGCGCGGGCCGAGGGCCCCGGGAGGCAGCGGCCCCGC
CCTCTTTACCTGCGGCCTCGCAGAGCATGCTGGGAGCCGCGGGAGGCAGTGGCCCCGCTCCCCTCACCTGCGGTATC
GCAGAGCATGGTGGGAGCCCCGGGAGGCAGTGGCCCCGCCCCCTTCCCTGCGGCCGC

<210> SEQ ID NO 119

<211> Length : 1,250

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 119

>R35137_PEA_1_PEA_1_PEA_1_T14

TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAGAGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAG
GCATGGACTGAGGGCGAAGGTGCTGACGCTGGACGGCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTG
GCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAGGAGCTGCGCCAGGGTGTGAAGAAGCCTTTCACCGAGGTCATC
CGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCACCTTCCTGCGCCAGGTCTTGGCCCTCTGTGT
TAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCGCATCTTGCAGGCGTGTG
GGGGCCACAGTCTGGGGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAG
AGGCGTGACGGAGGCATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTGACGGT
GCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACTCTACT
CGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGGACGTG
GCCGAGCTTCACCGTGCACTGGGCCAGGCGCGGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCGGATGAC
CATTCTGCCCCCCTTGGAGAAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCCTCGAGT
ACTCCTGAGCACCCCAGCTGGGGCCAGGCTGGGTCGCCCTGGACTGTGTGCTCAGGAGCCCTGGGAGGCTCTGGAGC

CCACTGTACTTGCTCTTGATGCCTGGCGGGGTGGGGTGGGGGGGGGTGCTGGGCCCCTGCCTCTCTGCAGGTCCCTAA
TAAAGCTGTGTGGCAGTC

<210> SEQ ID NO 120

<211> Length : 1,292

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 120

>Z25299_PEA_2_T1

GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTGGAGTCTGTCCT
CCTAAGAAATCTGCCCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAG
ATGTTGTCCTGACACTTGTGGCATCAAATGCCTGGATCCTGTTGACACCCCAAACCCAACAAGGAGGAAGCCTGGGA
AGTGCCCAGTGACTTATGGCCAATGTTTGATGCTTAACCCCCCCAATTTCTGTGAGATGGATGGCCAGTGCAAGCGT
GACTTGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTTTCCCCTGTGAAAGGTAAGCAGGGGATGAGGGCACA
CTGAGCTCCCTCCAGCCCTCTCAGCCTCAACCCTCTGGAGGCCCAGGCATATGGGCAGGGGGACTCCTGAACCCTAC
TCCAAGCACAGCCTCTGTCTGACTCCCTTGTCCTTCAAGAGAACTGTTCTCCAGGTCTCAGGGCCAGGATTTCCATA
GGATCGCCTGTGGCTTTGATTCTATTCTAGTGTCTCTGGGTGGGGGTCCTGGGCAAGTGTCTTTCTGAGTCTCAGTT
TCTTTATCGGTAAAATGTACATAATGAGATTGAAAGTGCTCTGCAAAGCACTATGTGCACTAAGAATTTATTATTCA
GGTTGTTTCCATCATGTTTTCTGAGGTGAAATCACAAAGGATCAGTGGAGTTTGAGGATTATCTAGTTCAATGCTTT
GAGTTTAGAGTTTTACGGTGAAAATGAGACTTGTCTCCTGACACTAAGTCTCTCTCAACTATAGCGCTATCTTGCTA
TTTTCTCTATCTCAGAAGGATCCTTGGGCAGGAGGAAGGATGTGGATATGATTTGGCTGGTTTCTATGCTGAAGCTC
TTATCTGATTTTCTCTCACAGCTTGATTCCTGCCATATGGAGGAGGCTCTGGAGTCCTGCTCTGTGTGGTCCAGGTC
CTTTCCACCCTGAGACTTGGCTCCACCACTGATATCCTCCTTTGGGGAAAGGCTTGGCACACAGCAGGCTTTCAAGA
AGTGCCAGTTGATCAATGAATAAATAAACGAGCCTATTTCTCTTTGCAAAACCTGCTTCT

<210> SEQ ID NO 121

<211> Length : 886

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 121

>Z25299_PEA_2_T2

GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTGGAGTCTGTCCT
CCTAAGAAATCTGCCCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAG
ATGTTGTCCTGACACTTGTGGCATCAAATGCCTGGATCCTGTTGACACCCCAAACCCAACAAGGAGGAAGCCTGGGA
AGTGCCCAGTGACTTATGGCCAATGTTTGATGCTTAACCCCCCCAATTTCTGTGAGATGGATGGCCAGTGCAAGCGT
GACTTGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTTTCCCCTGTGAAAGGTGAAAAGAGACATCACAAGCA
ATTGAGGGACCAGGAAGTGGATCCTCTAGAGATGAGGAGGCATTCTGCTGGATGACTTTTAAAAATGTTTTCTCCAG
AGTCATCTCTCTCATTAACAATGTTTTTTGTCTTAGAAATTTCTTGTTGATTTTTAAACTTACATGATTTCTTGTTT
TGGTATGAATACAGGCTGCTTCAGTCCTTCAATAAGCCCATCACACTTTTTCACCATGTCATCTATCAGCACTTTTT
CTGCAGTGTTACGAACATCAGCTTCATCACTGTCAGCCTGCGTTTTGCCTGCAACCCATCAAATGAGGTCAGGAGAG
GAGTTTTCCACTTTTGGCTTCATGTTGGTGCTCAAAACT

<210> SEQ ID NO 122

<211> Length : 696

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 122

>Z25299_PEA_2_T3

GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTGGAGTCTGTCCT
CCTAAGAAATCTGCCCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAG
ATGTTGTCCTGACACTTGTGGCATCAAATGCCTGGATCCTGTTGACACCCCAAACCCAACAAGGAGGAAGCCTGGGA
AGTGCCCAGTGACTTATGGCCAATGTTTGATGCTTAACCCCCCCAATTTCTGTGAGATGGATGGCCAGTGCAAGCGT
GACTTGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTTTCCCCTGTGAAAGGCTGCTTCAGTCCTTCAATAAG
CCCATCACACTTTTTCACCATGTCATCTATCAGCACTTTTTCTGCAGTGTTACGAACATCAGCTTCATCACTGTCAG
CCTGCGTTTTGCCTGCAACCCATCAAATGAGGTCAGGAGAGGAGTTTTCCACTTTTGGCTTCATGTTGGTGCTCAAA
ACT

<210> SEQ ID NO 123

<211> Length : 706

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 123

>Z25299_PEA_2_T6

GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTGGAGTCTGTCCT
CCTAAGAAATCTGCCCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAG
ATGTTGTCCTGACACTTGTGGCATCAAATGCCTGGATCCTGTTGACACCCCAAACCCAAGAGGAAGCCTGGGAAGTG
CCCAGTGACTTATGGCCAATGTTTGATGCTTAACCCCCCCAATTTCTGTGAGATGGATGGCCAGTGCAAGCGTGACT
TGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTTTCCCCTGTGAAAGCTTGATTCCTGCCATATGGAGGAGGC
TCTGGAGTCCTGCTCTGTGTGGTCCAGGTCCTTTCCACCCTGAGACTTGGCTCCACCACTGATATCCTCCTTTGGGG
AAAGGCTTGGCACACAGCAGGCTTTCAAGAAGTGCCAGTTGATCAATGAATAAATAAACGAGCCTATTTCTCTTTGC
AAAACCTGCTTCT

<210> SEQ ID NO 124

<211> Length : 560

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 124

>Z25299_PEA_2_T9

GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGTCCTTCAAAGCTGGAGTCTGTCCT
CCTAAGAAATCTGCCCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAG
ATGTTGTCCTGACACTTGTGGCATCAAATGCCTGGATCCTGTTGACACCCCAAACCCAACTTGATTCCTGCCATATG
GAGGAGGCTCTGGAGTCCTGCTCTGTGTGGTCCAGGTCCTTTCCACCCTGAGACTTGGCTCCACCACTGATATCCTC
CTTTGGGGAAAGGCTTGGCACACAGCAGGCTTTCAAGAAGTGCCAGTTGATCAATGAATAAATAAACGAGCCTATTT
CTCTTTGCAAAACCTGCTTCT

<210> SEQ ID NO 125

<211> Length : 3,194

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 125

>HSSTROL3_T5

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAGGATCCTTCGGTTCCCATGGCAGTT
GGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCGATGTGACGCCACTCACCTTTACTG
AGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACTGGCATGGGGACGACCTGCCGTTTGATGGG
CCTGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAAGGGGATGTCCACTTCGACTATGATGAGAC
CTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCACGTGCTGGGGCTGC
AGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGAC
TGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGC
TGGGATAGACACCAATGAGATTGCACCGCTGGAGCCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGG
TCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGTGGCGCCTCCGTGGGGGCCAGCTGCAGCCC
GGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCTGTGGACGCTGCCTTCGAGGATGCCCAGGG
CCACATTTGGTTCTTCCAAGGTGCTCAGTACTGGGTGTACGACGGTGAAAAGCCAGTCCTGGGCCCCGCACCCCTCA
CCGAGCTGGGCCTGGTGAGGTTCCCGGTCCATGCTGCCTTGGTCTGGGGTCCCGAGAAGAACAAGATCTACTTCTTC
CGAGGCAGGGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAGGGCCACTGACTG
GAGAGGGGTGCCCTCTGAGATCGACGCTGCCTTCCAGGATGCTGATGGTGCGTTGGGGGTGAGGCAGCTGGTGGGAG
GTGGGCACAGCAGCCGCTTCTCCCACCTGGTGGTGGCTGGGCTCCCACATGCCTGCCACAGGAAGTCTGGCTCTTCA
TCACAGGTCCTTTGTCCAGAGCCATCTGCCCTCCTCTCGGTGGCCGGCTAGTGCTACATTCCATATTGCAGATGAGG
AAACTGAGGGTCAGAGAAGTGCAAGGTCTTACCCTGGTTTTTCAGCCACAGCCAGTAGAACAATAAACTGCTGTACA
CTGAGGGCCAACAATGCTCTAAGCTCCTTACTGGTCTCATCCAGTTCTCAGAACAGCCCTCTGATGTGACACCTGTT
GTGAACCCAGTTTCCAGAGGAGCAAACAGAGGCTCAGGCAATGAGGCCCCTAACCTGGACTACCCTGGTGGTCCCTG
CTCCTAACCACTGACCCACCCAGCCTCCCACAACCACAGGGGGCTAGAGCCAGTCCAGTGCTCCCTCCCCTGCTAGG
CTCCTCTTCTGTGCTCTTTCTCCCACATCAGGACCCACTGGGAGAGCTATCCTAGGGTAGCCTCCAGCTCCAGGACT
CCAGGGTGCCCGTCAATAGCCTGGCTAATTTAATAGATGCAGGAGAGAGTGATGTGGAGGGTGGTGGGGGCAACGGG
ACTTGCTTTCCTGAGAGGTGGGACTCAGGCCTCTGAGGCTCTGGGTACCTGTCAGGCTGGGTATTAGCCCAGCCCAG
ATTCCGGGGCAGGCAGAAGGGCTCCCTAGAGGGAAGAGAGGTTCTGAAAGGCCGGCCCTGGATCCTGCAGGACTCGA
GGAACTCAGCAGTGGCCAAGGGCTTCCCACTCAGCCCTCCCTTAGTGCCCATCCCTGGGCACAGCCTGACAGGCAGG
AGTAGGGCCCAGTGTCCACTCGCCCAGGCTTGACCACCTTCTCTTCTCAGGCTATGCCTACTTCCTGCGCGGCCGCC
TCTACTGGAAGTTTGACCCTGTGAAGGTGAAGGCTCTGGAAGGCTTCCCCCGTCTCGTGGGTCCTGACTTCTTTGGC
TGTGCCGAGCCTGCCAACACTTTCCTCTGACCATGGCTTGGATGCCCTCAGGGGTGCTGACCCCTGCCAGGCCACGA

ATATCAGGCTAGAGACCCATGGCCATCTTTGTGGCTGTGGGCACCAGGCATGGGACTGAGCCCATGTCTCCTCAGGG
GGATGGGGTGGGGTACAACCACCATGACAACTGCCGGGAGGGCCACGCAGGTCGTGGTCACCTGCCAGCGACTGTCT
CAGACTGGGCAGGGAGGCTTTGGCATGACTTAAGAGGAAGGGCAGTCTTGGGCCCGCTATGCAGGTCCTGGCAAACC
TGGCTGCCCTGTCTCCATCCCTGTCCCTCAGGGTAGCACCATGGCAGGACTGGGGGAACTGGAGTGTCCTTGCTGTA
TCCCTGTTGTGAGGTTCCTTCCAGGGGCTGGCACTGAAGCAAGGGTGCTGGGGCCCCATGGCCTTCAGCCCTGGCTG
AGCAACTGGGCTGTAGGGCAGGGCCACTTCCTGAGGTCAGGTCTTGGTAGGTGCCTGCATCTGTCTGCCTTCTGGCT
GACAATCCTGGAAATCTGTTCTCCAGAATCCAGGCCAAAAAGTTCACAGTCAAATGGGGAGGGGTATTCTTCATGCA
GGAGACCCCAGGCCCTGGAGGCTGCAACATACCTCAATCCTGTCCCAGGCCGGATCCTCCTGAAGCCCTTTTCGCAG
CACTGCTATCCTCCAAAGCCATTGTAAATGTGTGTACAGTGTGTATAAACCTTCTTCTTCTTTTTTTTTTTTAAACT
GAGGATTGTCATTAAACACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 126

<211> Length : 2,705

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 126

>HSSTROL3_T8

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAGGATCCTTCGGTTCCCATGGCAGTT
GGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCGATGTGACGCCACTCACCTTTACTG
AGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACTGGCATGGGGACGACCTGCCGTTTGATGGG
CCTGGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAAGGGGATGTCCACTTCGACTATGATGAGAC
CTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCACGTGCTGGGGCTGC
AGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGAC
TGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGC
TGGGATAGACACCAATGAGATTGCACCGCTGGAGCCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGG
TCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGTGGCGCCTCCGTGGGGGCCAGCTGCAGCCC
GGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCTGTGGACGCTGCCTTCGAGGATGCCCAGGG
CCACATTTGGTTCTTCCAAGAGCTGGGATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCC
AGGGACTTCACAAATGAAGGCACAGCATGGGAAACCTGCGTGGGTTCCAGGGCAGTCCAGCCTGCAGGGGCCCAGGG
AGTGGTCAGTAGGCATTTGTCACAGCCAAATGCCAGTGGAAGGAGCAGCCGCCCAGGCAGCCCTCTACTGATGAGAG
TAACCTCACCCGTGCACTAGTTTACAGAGCATTCACTGCCCCAGCTTATCCCAGGCCTCCCGCTTCCCTCTGCGGGT

GGGGTGCTGAGCAGGCATTATTGGCCTGCATGTTTTACTGATGAGGAAACTGAGGCTGGGAGAGTCTGTGGTAGGGG
TCAAGCAGGTCCACAGTGGCGGGGCATGGCAGTGGTGGCTGGGCAGGTCCTTGCAGCCTTCCCTCTCCGGCAGGTGC
TCAGTACTGGGTGTACGACGGTGAAAAGCCAGTCCTGGGCCCCGCACCCCTCACCGAGCTGGGCCTGGTGAGGTTCC
CGGTCCATGCTGCCTTGGTCTGGGGTCCCGAGAAGAACAAGATCTACTTCTTCCGAGGCAGGGACTACTGGCGTTTC
CACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAGGGCCACTGACTGGAGAGGGGTGCCCTCTGAGATCGA
CGCTGCCTTCCAGGATGCTGATGGCTATGCCTACTTCCTGCGCGGCCGCCTCTACTGGAAGTTTGACCCTGTGAAGG
TGAAGGCTCTGGAAGGCTTCCCCCGTCTCGTGGGTCCTGACTTCTTTGGCTGTGCCGAGCCTGCCAACACTTTCCTC
TGACCATGGCTTGGATGCCCTCAGGGGTGCTGACCCCTGCCAGGCCACGAATATCAGGCTAGAGACCCATGGCCATC
TTTGTGGCTGTGGGCACCAGGCATGGGACTGAGCCCATGTCTCCTCAGGGGGATGGGGTGGGGTACAACCACCATGA
CAACTGCCGGGAGGGCCACGCAGGTCGTGGTCACCTGCCAGCGACTGTCTCAGACTGGGCAGGGAGGCTTTGGCATG
ACTTAAGAGGAAGGGCAGTCTTGGGCCCGCTATGCAGGTCCTGGCAAACCTGGCTGCCCTGTCTCCATCCCTGTCCC
TCAGGGTAGCACCATGGCAGGACTGGGGGAACTGGAGTGTCCTTGCTGTATCCCTGTTGTGAGGTTCCTTCCAGGGG
CTGGCACTGAAGCAAGGGTGCTGGGGCCCCATGGCCTTCAGCCCTGGCTGAGCAACTGGGCTGTAGGGCAGGGCCAC
TTCCTGAGGTCAGGTCTTGGTAGGTGCCTGCATCTGTCTGCCTTCTGGCTGACAATCCTGGAAATCTGTTCTCCAGA
ATCCAGGCCAAAAAGTTCACAGTCAAATGGGGAGGGGTATTCTTCATGCAGGAGACCCCAGGCCCTGGAGGCTGCAA
CATACCTCAATCCTGTCCCAGGCCGGATCCTCCTGAAGCCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTAA
ATGTGTGTACAGTGTGTATAAACCTTCTTCTTCTTTTTTTTTTTTAAACTGAGGATTGTCATTAAACACAGTTGTTT
TCTACCTGCC

<210> SEQ ID NO 127

<211> Length : 3,32

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 127

>HSSTROL3_T9

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAGGATCCTTCGGTTCCCATGGCAGTT
GGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCGATGTGACGCCACTCACCTTTACTG
AGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACTGGCATGGGGACGACCTGCCGTTTGATGGG
CCTGGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAAGGGGATGTCCACTTCGACTATGATGAGAC
CTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCACGTGCTGGGGCTGC
AGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGAC

TGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGC
TGGGATAGACACCAATGAGATTGCACCGCTGGAGCCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGG
TCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGTGGCGCCTCCGTGGGGGCCAGCTGCAGCCC
GGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCTGTGGACGCTGCCTTCGAGGATGCCCAGGG
CCACATTTGGTTCTTCCAAGAGCTGGGATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCC
AGGGACTTCACAAATGAAGGCACAGCATGGGAAACCTGCGTGGGTTCCAGGGCAGTCCAGCCTGCAGGGGCCCAGGG
AGTGGTGCTCAGTACTGGGTGTACGACGGTGAAAAGCCAGTCCTGGGCCCCGCACCCCTCACCGAGCTGGGCCTGGT
GAGGTTCCCGGTCCATGCTGCCTTGGTCTGGGGTCCCGAGAAGAACAAGATCTACTTCTTCCGAGGCAGGGACTACT
GGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAGGGCCACTGACTGGAGAGGGGTGCCCTCT
GAGATCGACGCTGCCTTCCAGGATGCTGATGGTGCGTTGGGGGTGAGGCAGCTGGTGGGAGGTGGGCACAGCAGCCG
CTTCTCCCACCTGGTGGTGGCTGGGCTCCCACATGCCTGCCACAGGAAGTCTGGCTCTTCATCACAGGTCCTTTGTC
CAGAGCCATCTGCCCTCCTCTCGGTGGCCGGCTAGTGCTACATTCCATATTGCAGATGAGGAAACTGAGGGTCAGAG
AAGTGCAAGGTCTTACCCTGGTTTTTCAGCCACAGCCAGTAGAACAATAAACTGCTGTACACTGAGGGCCAACAATG
CTCTAAGCTCCTTACTGGTCTCATCCAGTTCTCAGAACAGCCCTCTGATGTGACACCTGTTGTGAACCCAGTTTCCA
GAGGAGCAAACAGAGGCTCAGGCAATGAGGCCCCTAACCTGGACTACCCTGGTGGTCCCTGCTCCTAACCACTGACC
CACCCAGCCTCCCACAACCACAGGGGGCTAGAGCCAGTCCAGTGCTCCCTCCCCTGCTAGGCTCCTCTTCTGTGCTC
TTTCTCCCACATCAGGACCCACTGGGAGAGCTATCCTAGGGTAGCCTCCAGCTCCAGGACTCCAGGGTGCCCGTCAA
TAGCCTGGCTAATTTAATAGATGCAGGAGAGAGTGATGTGGAGGGTGGTGGGGGCAACGGGACTTGCTTTCCTGAGA
GGTGGGACTCAGGCCTCTGAGGCTCTGGGTACCTGTCAGGCTGGGTATTAGCCCAGCCCAGATTCCGGGGCAGGCAG
AAGGGCTCCCTAGAGGGAAGAGAGGTTCTGAAAGGCCGGCCCTGGATCCTGCAGGACTCGAGGAACTCAGCAGTGGC
CAAGGGCTTCCCACTCAGCCCTCCCTTAGTGCCCATCCCTGGGCACAGCCTGACAGGCAGGAGTAGGGCCCAGTGTC
CACTCGCCCAGGCTTGACCACCTTCTCTTCTCAGGCTATGCCTACTTCCTGCGCGGCCGCCTCTACTGGAAGTTTGA
CCCTGTGAAGGTGAAGGCTCTGGAAGGCTTCCCCCGTCTCGTGGGTCCTGACTTCTTTGGCTGTGCCGAGCCTGCCA
ACACTTTCCTCTGACCATGGCTTGGATGCCCTCAGGGGTGCTGACCCCTGCCAGGCCACGAATATCAGGCTAGAGAC
CCATGGCCATCTTTGTGGCTGTGGGCACCAGGCATGGGACTGAGCCCATGTCTCCTCAGGGGGATGGGGTGGGGTAC
AACCACCATGACAACTGCCGGGAGGGCCACGCAGGTCGTGGTCACCTGCCAGCGACTGTCTCAGACTGGGCAGGGAG
GCTTTGGCATGACTTAAGAGGAAGGGCAGTCTTGGGCCCGCTATGCAGGTCCTGGCAAACCTGGCTGCCCTGTCTCC
ATCCCTGTCCCTCAGGGTAGCACCATGGCAGGACTGGGGGAACTGGAGTGTCCTTGCTGTATCCCTGTTGTGAGGTT
CCTTCCAGGGGCTGGCACTGAAGCAAGGGTGCTGGGGCCCCATGGCCTTCAGCCCTGGCTGAGCAACTGGGCTGTAG
GGCAGGGCCACTTCCTGAGGTCAGGTCTTGGTAGGTGCCTGCATCTGTCTGCCTTCTGGCTGACAATCCTGGAAATC
TGTTCTCCAGAATCCAGGCCAAAAAGTTCACAGTCAAATGGGGAGGGGTATTCTTCATGCAGGAGACCCCAGGCCCT
GGAGGCTGCAACATACCTCAATCCTGTCCCAGGCCGGATCCTCCTGAAGCCCTTTTCGCAGCACTGCTATCCTCCAA
AGCCATTGTAAATGTGTGTACAGTGTGTATAAACCTTCTTCTTCTTTTTTTTTTTTAAACTGAGGATTGTCATTAAA
CACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 128

<211> Length : 3,052

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 128

>HSSTROL3_T10

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAGGATCCTTCGGTTCCCATGGCAGTT
GGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCGATGTGACGCCACTCACCTTTACTG
AGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACTGGCATGGGGACGACCTGCCGTTTGATGGG
CCTGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAAGGGGATGTCCACTTCGACTATGATGAGAC
CTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCACGTGCTGGGGCTGC
AGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGAC
TGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGC
TGGGATAGACACCAATGAGATTGCACCGCTGGAGCCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGG
TCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGTGGCGCCTCCGTGGGGGCCAGCTGCAGCCC
GGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCTGTGGACGCTGCCTTCGAGGATGCCCAGGG
CCACATTTGGTTCTTCCAAGGGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAG
GGCCACTGACTGGAGAGGGGTGCCCTCTGAGATCGACGCTGCCTTCCAGGATGCTGATGGTGCGTTGGGGGTGAGGC
AGCTGGTGGGAGGTGGGCACAGCAGCCGCTTCTCCCACCTGGTGGTGGCTGGGCTCCCACATGCCTGCCACAGGAAG
TCTGGCTCTTCATCACAGGTCCTTTGTCCAGAGCCATCTGCCCTCCTCTCGGTGGCCGGCTAGTGCTACATTCCATA
TTGCAGATGAGGAAACTGAGGGTCAGAGAAGTGCAAGGTCTTACCCTGGTTTTTCAGCCACAGCCAGTAGAACAATA
AACTGCTGTACACTGAGGGCCAACAATGCTCTAAGCTCCTTACTGGTCTCATCCAGTTCTCAGAACAGCCCTCTGAT
GTGACACCTGTTGTGAACCCAGTTTCCAGAGGAGCAAACAGAGGCTCAGGCAATGAGGCCCCTAACCTGGACTACCC
TGGTGGTCCCTGCTCCTAACCACTGACCCACCCAGCCTCCCACAACCACAGGGGGCTAGAGCCAGTCCAGTGCTCCC
TCCCCTGCTAGGCTCCTCTTCTGTGCTCTTTCTCCCACATCAGGACCCACTGGGAGAGCTATCCTAGGGTAGCCTCC
AGCTCCAGGACTCCAGGGTGCCCGTCAATAGCCTGGCTAATTTAATAGATGCAGGAGAGAGTGATGTGGAGGGTGGT
GGGGGCAACGGGACTTGCTTTCCTGAGAGGTGGGACTCAGGCCTCTGAGGCTCTGGGTACCTGTCAGGCTGGGTATT
AGCCCAGCCCAGATTCCGGGGCAGGCAGAAGGGCTCCCTAGAGGGAAGAGAGGTTCTGAAAGGCCGGCCCTGGATCC
TGCAGGACTCGAGGAACTCAGCAGTGGCCAAGGGCTTCCCACTCAGCCCTCCCTTAGTGCCCATCCCTGGGCACAGC
CTGACAGGCAGGAGTAGGGCCCAGTGTCCACTCGCCCAGGCTTGACCACCTTCTCTTCTCAGGCTATGCCTACTTCC
TGCGCGGCCGCCTCTACTGGAAGTTTGACCCTGTGAAGGTGAAGGCTCTGGAAGGCTTCCCCCGTCTCGTGGGTCCT
GACTTCTTTGGCTGTGCCGAGCCTGCCAACACTTTCCTCTGACCATGGCTTGGATGCCCTCAGGGGTGCTGACCCCT
GCCAGGCCACGAATATCAGGCTAGAGACCCATGGCCATCTTTGTGGCTGTGGGCACCAGGCATGGGACTGAGCCCAT
GTCTCCTCAGGGGGATGGGGTGGGGTACAACCACCATGACAACTGCCGGGAGGGCCACGCAGGTCGTGGTCACCTGC

CAGCGACTGTCTCAGACTGGGCAGGGAGGCTTTGGCATGACTTAAGAGGAAGGGCAGTCTTGGGCCCGCTATGCAGG
TCCTGGCAAACCTGGCTGCCCTGTCTCCATCCCTGTCCCTCAGGGTAGCACCATGGCAGGACTGGGGGAACTGGAGT
GTCCTTGCTGTATCCCTGTTGTGAGGTTCCTTCCAGGGGCTGGCACTGAAGCAAGGGTGCTGGGGCCCCATGGCCTT
CAGCCCTGGCTGAGCAACTGGGCTGTAGGGCAGGGCCACTTCCTGAGGTCAGGTCTTGGTAGGTGCCTGCATCTGTC
TGCCTTCTGGCTGACAATCCTGGAAATCTGTTCTCCAGAATCCAGGCCAAAAAGTTCACAGTCAAATGGGGAGGGGT
ATTCTTCATGCAGGAGACCCCAGGCCCTGGAGGCTGCAACATACCTCAATCCTGTCCCAGGCCGGATCCTCCTGAAG
CCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTAAATGTGTGTACAGTGTGTATAAACCTTCTTCTTCTTTTT
TTTTTTTAAACTGAGGATTGTCATTAAACACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 129

<211> Length : 2,359

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 129

>HSSTROL3_T11

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAGGATCCTTCGGTTCCCATGGCAGTT
GGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCGATGTGACGCCACTCACCTTTACTG
AGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACTGGCATGGGGACGACCTGCCGTTTGATGGG
CCTGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAAGGGGATGTCCACTTCGACTATGATGAGAC
CTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCACGTGCTGGGGCTGC
AGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGAC
TGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGC
TGGGATAGACACCAATGAGATTGCACCGCTGGAGGTGAGGCCCTGCCTGCCAGTCCCCCTACTCCTCTGCTGGCCAC
TGTGACTGCAGCATATGCCCTCAGCATGTGTCCCTCTCTCCCACCCCAGCCAGACGCCCCGCCAGATGCCTGTGAGG
CCTCCTTTGACGCGGTCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGTGGCGCCTCCGTGGG
GGCCAGCTGCAGCCCGGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCTGTGGACGCTGCCTT
CGAGGATGCCCAGGGCCACATTTGGTTCTTCCAAGGTGCTCAGTACTGGGTGTACGACGGTGAAAAGCCAGTCCTGG
GCCCCGCACCCCTCACCGAGCTGGGCCTGGTGAGGTTCCCGGTCCATGCTGCCTTGGTCTGGGGTCCCGAGAAGAAC
AAGATCTACTTCTTCCGAGGCAGGGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCG
CAGGGCCACTGACTGGAGAGGGGTGCCCTCTGAGATCGACGCTGCCTTCCAGGATGCTGATGGCTATGCCTACTTCC
TGCGCGGCCGCCTCTACTGGAAGTTTGACCCTGTGAAGGTGAAGGCTCTGGAAGGCTTCCCCCGTCTCGTGGGTCCT

GACTTCTTTGGCTGTGCCGAGCCTGCCAACACTTTCCTCTGACCATGGCTTGGATGCCCTCAGGGGTGCTGACCCCT
GCCAGGCCACGAATATCAGGCTAGAGACCCATGGCCATCTTTGTGGCTGTGGGCACCAGGCATGGGACTGAGCCCAT
GTCTCCTCAGGGGGATGGGGTGGGGTACAACCACCATGACAACTGCCGGGAGGGCCACGCAGGTCGTGGTCACCTGC
CAGCGACTGTCTCAGACTGGGCAGGGAGGCTTTGGCATGACTTAAGAGGAAGGGCAGTCTTGGGCCCGCTATGCAGG
TCCTGGCAAACCTGGCTGCCCTGTCTCCATCCCTGTCCCTCAGGGTAGCACCATGGCAGGACTGGGGGAACTGGAGT
GTCCTTGCTGTATCCCTGTTGTGAGGTTCCTTCCAGGGGCTGGCACTGAAGCAAGGGTGCTGGGGCCCCATGGCCTT
CAGCCCTGGCTGAGCAACTGGGCTGTAGGGCAGGGCCACTTCCTGAGGTCAGGTCTTGGTAGGTGCCTGCATCTGTC
TGCCTTCTGGCTGACAATCCTGGAAATCTGTTCTCCAGAATCCAGGCCAAAAAGTTCACAGTCAAATGGGGAGGGGT
ATTCTTCATGCAGGAGACCCCAGGCCCTGGAGGCTGCAACATACCTCAATCCTGTCCCAGGCCGGATCCTCCTGAAG
CCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTAAATGTGTGTACAGTGTGTATAAACCTTCTTCTTCTTTTT
TTTTTTTAAACTGAGGATTGTCATTAAACACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 130

<211> Length : 2,077

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 130

>HSSTROL3_T12

CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCGGACGTCCACCACCTCCATGCCGA
GAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGCCCCTGCCACGCAGGAAGCCCCCC
GGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATGGGCTGAGTGCCCGCAACCGACAG
AAGAGGATCCTTCGGTTCCCATGGCAGTTGGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATG
GAGCGATGTGACGCCACTCACCTTTACTGAGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAGGTACT
GGCATGGGGACGACCTGCCGTTTGATGGGCCTGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACCGAGAA
GGGGATGTCCACTTCGACTATGATGAGACCTGGACTATCGGGGATGACCAGGGCACAGACCTGCTGCAGGTGGCAGC
CCATGAATTTGGCCACGTGCTGGGGCTGCAGCACACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTC
GCTACCCACTGAGTCTCAGCCCAGATGACTGCAGGGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACC
TCCAGGACCCCAGCCCTGGGCCCCCAGGCTGGGATAGACACCAATGAGATTGCACCGCTGGAGCCAGACGCCCCGCC
AGATGCCTGTGAGGCCTCCTTTGACGCGGTCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAAAGCGGGCTTTGTGT
GGCGCCTCCGTGGGGGCCAGCTGCAGCCCGGCTACCCAGCATTGGCCTCTCGCCACTGGCAGGGACTGCCCAGCCCT
GTGGACGCTGCCTTCGAGGATGCCCAGGGCCACATTTGGTTCTTCCAAGGGACTACTGGCGTTTCCACCCCAGCACC
CGGCGTGTAGACAGTCCCGTGCCCCGCAGGGCCACTGACTGGAGAGGGGTGCCCTCTGAGATCGACGCTGCCTTCCA
GGATGCTGATGGCTATGCCTACTTCCTGCGCGGCCGCCTCTACTGGAAGTTTGACCCTGTGAAGGTGAAGGCTCTGG
AAGGCTTCCCCCGTCTCGTGGGTCCTGACTTCTTTGGCTGTGCCGAGCCTGCCAACACTTTCCTCTGACCATGGCTT

GGATGCCCTCAGGGGTGCTGACCCCTGCCAGGCCACGAATATCAGGCTAGAGACCCATGGCCATCTTTGTGGCTGTG
GGCACCAGGCATGGGACTGAGCCCATGTCTCCTCAGGGGGATGGGGTGGGGTACAACCACCATGACAACTGCCGGGA
GGGCCACGCAGGTCGTGGTCACCTGCCAGCGACTGTCTCAGACTGGGCAGGGAGGCTTTGGCATGACTTAAGAGGAA
GGGCAGTCTTGGGCCCGCTATGCAGGTCCTGGCAAACCTGGCTGCCCTGTCTCCATCCCTGTCCCTCAGGGTAGCAC
CATGGCAGGACTGGGGGAACTGGAGTGTCCTTGCTGTATCCCTGTTGTGAGGTTCCTTCCAGGGGCTGGCACTGAAG
CAAGGGTGCTGGGGCCCCATGGCCTTCAGCCCTGGCTGAGCAACTGGGCTGTAGGGCAGGGCCACTTCCTGAGGTCA
GGTCTTGGTAGGTGCCTGCATCTGTCTGCCTTCTGGCTGACAATCCTGGAAATCTGTTCTCCAGAATCCAGGCCAAA
AAGTTCACAGTCAAATGGGGAGGGGTATTCTTCATGCAGGAGACCCCAGGCCCTGGAGGCTGCAACATACCTCAATC
CTGTCCCAGGCCGGATCCTCCTGAAGCCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTAAATGTGTGTACAG
TGTGTATAAACCTTCTTCTTCTTTTTTTTTTTAAACTGAGGATTGTCATTAAACACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 131

<211> Length : 1,266

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 131

>HUMTREFAC_PEA_2_T4

CGCTCCCCAGTAGAGGACCCGGAACCAGAACTGGAATCCGCCCTTACCGCTTGCTGCCAAAACAGTGGGGGCTGAAC
TGACCTCTCCCCTTTGGGAGAGAAAAACTGTCTGGGAGCTTGACAAAGGCATGCAGGAGAGAACAGGAGCAGCCACA
GCCAGGAGGGAGAGCCTTCCCCAAGCAAACAATCCAGAGCAGCTGTGCAAACAACGGTGCATAAATGAGGCCTCCTG
GACCATGAAGCGAGTCCTGAGCTGCGTCCCGGAGCCCACGGTGGTCATGGCTGCCAGAGCGCTCTGCATGCTGGGGC
TGGTCCTGGCCTTGCTGTCCTCCAGCTCTGCTGAGGAGTACGTGGGCCTGTGGAAAGTGCATCTTCCTAAGGGCGAG
GGTTTCAGCAGTGGTTGAACTCGGCGGGGTGGGGCGGAGCGGGAGGATGCAAACTTGCAAAGTGAAGCAAACACACT
CACCGCAGCCCAGCAAGGGCTCTGGCAGCTGACAGGGCTTTGTCTGGGACAGCTGCAAACCAGTGTGCCGTGCCAGC
CAAGGACAGGGTGGACTGCGGCTACCCCCATGTCACCCCCAAGGAGTGCAACAACCGGGGCTGCTGCTTTGACTCCA
GGATCCCTGGAGTGCCTTGGTGTTTCAAGCCCCTGCAGGAAGCAGAATGCACCTTCTGAGGCACCTCCAGCTGCCCC
CGGCCGGGGGATGCGAGGCTCGGAGCACCCTTGCCCGGCTGTGATTGCTGCCAGGCACTGTTCATCTCAGCTTTTCT
GTCCCTTTGCTCCCGGCAAGCGCTTCTGCTGAAAGTTCATATCTGGAGCCTGATGTCTTAACGAATAAAGGTCCCAT
GCTCCACCCGAGGACAGTTCTTCGTGCCTGAGACTTTCTGAGGTTGTGCTTTATTTCTGCTGCGTCGTGGGAGAGGG
CGGGAGGGTGTCAGGGGAGAGTCTGCCCAGGCCTCAAGGGCAGGAAAAGACTCCCTAAGGAGCTGCAGTGCATGCAA
GGATATTTTGAATCCAGACTGGCACCCACGTCACAGGAAAGCCTAGGAACACTGTAAGTGCCGCTTCCTCGGGAAAG
CAGAAAAAATACATTTCAGGTAGAAGTTTTCAAAAATCACAAGTCTTTCTTGGTGAAGACAGCAAGCCAATAAAACT

GTCTTCCAAAGTGGTCCTTTATTTCACAACCACTCTCGCTACTGTTCAATACTTGTACTATTCCTGGGTTTTGTTTC
TTTGTACAGTAAACATTATGAACAAACAGGCAAA

<210> SEQ ID NO 132

<211> Length : 747

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 132
>HUMTREFAC_PEA_2_T5
CGCTCCCCAGTAGAGGACCCGGAACCAGAACTGGAATCCGCCCTTACCGCTTGCTGCCAAAACAGTGGGGGCTGAAC
TGACCTCTCCCCTTTGGGAGAGAAAAACTGTCTGGGAGCTTGACAAAGGCATGCAGGAGAGAACAGGAGCAGCCACA
GCCAGGAGGGAGAGCCTTCCCCAAGCAAACAATCCAGAGCAGCTGTGCAAACAACGGTGCATAAATGAGGCCTCCTG
GACCATGAAGCGAGTCCTGAGCTGCGTCCCGGAGCCCACGGTGGTCATGGCTGCCAGAGCGCTCTGCATGCTGGGGC
TGGTCCTGGCCTTGCTGTCCTCCAGCTCTGCTGAGGAGTACGTGGGCCTGTCCCAGCAAGGGCTCTGGCAGCTGACA
GGGCTTTGTCTGGGACAGCTGCAAACCAGTGTGCCGTGCCAGCCAAGGACAGGGTGGACTGCGGCTACCCCCATGTC
ACCCCCAAGGAGTGCAACAACCGGGGCTGCTGCTTTGACTCCAGGATCCCTGGAGTGCCTTGGTGTTTCAAGCCCCT
GCAGGAAGCAGAATGCACCTTCTGAGGCACCTCCAGCTGCCCCCGGCCGGGGGATGCGAGGCTCGGAGCACCCTTGC
CCGGCTGTGATTGCTGCCAGGCACTGTTCATCTCAGCTTTTCTGTCCCTTTGCTCCCGGCAAGCGCTTCTGCTGAAA
GTTCATATCTGGAGCCTGATGTCTTAACGAATAAAGGTCCCATGCTCCACCCGA

<210> SEQ ID NO 133

<211> Length : 2,201

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 133
>HSS100PCB_T1
TGAGACAAGATGTCACTCTGTCACCCAGGCTGGAGTGCAGTGGCAGGATCACGGCTCACTGCAGCCTCGACCTCCCT
GGGCTCAGGTGATCCTCCCACCTCAGCCTACCGAGTAGCTGGGACTACAGGTGCATGTCACCATACCCGGTTAATTT
TTGTATTTTTTTTAGAGACAAGGTCTCACCATGTTGCCCAGGCTGGTCTCAAACTCCTGTGCTCAGGCAATGCGTCA
GCCTCGACATCTCAAAGTGCTGTGATTACAGGCGTGAGCCCCGACACCTGGCCTAGTTCTATTTTCTAAATGTGAAT
TCTGTAAAGATATCTTTTAAAAATAAAGTTCTGTTTTTGGTAGAAAATGTAAAAATAGATAAATATGGAGGGAAGAA

ATCCCCCCTGGAATACAGACGCTTCCTCTCCCTTCCAGCCTTTTCCCCATATGAACATTGCTGTGAGTGAGACTTAC
ATGCAATGTAATTTCTTTTTGAGCTTAACATTACAACATAAATTCTCAAACTCTGATGTTCATTAAACACCCCAGCC
CCATCCTGGGAACTTGGGCTTGGGGCTCGGGGTGTTCTGATAATGATCAAAGTATGAGAATTGAACCCATGAGGACT
TTGATCCAAGATACTGGGGTGTGGGGAGGGGCAGGCACAGGTGTCCTGGGAACACACTTTGAGAAGCAATGGCAAAG
CTGGGGGTCCAGCTAATGTGTTACATTAGAATCACCTCGGGGAGGCCCTGGGTGCCCTTCTCAGCCCTCCCTCCGGA
GGCTGCTGAAGCCCAGCAAAGCCGGAGTCAGAGAACAATGTCCGCCTGAGGGCAGGGCTGGGCTGGGCTGGCCTTCT
GGCCCTATCTGCTCCGTGCCCAACCCAGCGCCCCGCACAGTCGGAGCTTTGTAAATACGAGGTGACTGTCTGCCTAC
AAACTTTGTAAACATCACTTGAAATGGCCGCAGGGCATTGCGACATGGCCATACCACTATTTGTTTGCTATTGAATT
TGTACTTCCCTGCCTTACTTTTGCTATTGCAAACCATGCTGTCACTAAGGTCTTCATGCACACAGTTGTGTCTTGGT
CAGATGATATGTTTCTACCAATTTTAATTGTGTTTCTTTCCACCTGGACACACAGCTCTCTGGCCCAGGGCTGGGTC
ATCAGCACACCCTGCTGCTGCTGTTCAGATCTGCATCCTGGTCCCGCTTGGTCCCACAGTGAGAACGCTTTGCTATC
ACATGGGCAGGCTCTGAGAGCCCTGCCGGCCTGGCCTTCTCAAAGAAGACCTGAGAGCTTGGGACCCAAGCAGAGAG
GAAGAACAGGGCTCAGGGTGCTTGCTCCATGCTCGCTCCACACCTGGGGCTCAACCCTGGCTTTCCCCGGCTCCCTG
TGTGACTTCAGGGCAGGTCCCTTGGGCCCTCTGGGCCTTATCATCTTCATCTGTAACAGGGCGATGCCTCTGCCGTG
TCTGGTGGTGTTGAGGAGTTCCTGTTTGTGTAAGCAGCTAGTTCAGTGCCAGCACGAGATGGGAGGCCCATGAAGTT
AGCAGTGCACAAAAAATAGAGCAAAGACTGGATGCATTTCCTGAGAACAACCATCACTGTAAAGCACTTTACAAATC
CAAAGACAACCCCCGGCAAAAACTCAAAATGAAACTCCCTCTCGCAGAGCACAATTCCAATTCGCTCTAAAAACATT
ACAAGTTAGTTCATGTCATGCCAGATAGCTGAAGGCAGCTCACAAGTTCTTAAGGCCAGGAATGCCATGTGTCTGCT
ATGCACAGCTGGCCCTGGCCCTGAGCCTGAATGACAGCACAAAGGTGACGCAGATGTGGGTGCCCTGCTCCTGCCCA
GCAGCAGTGCTTGGTGGAGGCTGAGGCCCTGCACAGGCACCCTCACTGCTGACCTTGAGCCTCTCTCTCCTCTAGAG
TGGAAAAGACAAGGATGCCGTGGATAAATTGCTCAAGGACCTGGACGCCAATGGAGATGCCCAGGTGGACTTCAGTG
AGTTCATCGTGTTCGTGGCTGCAATCACGTCTGCCTGTCACAAGTACTTTGAGAAGGCAGGACTCAAATGATGCCCT
GGAGATGTCACAGATTCCTGGCAGAGCCATGGTCCCAGGCTTCCCAAAAGTGTTTGTTGGCAATTATTCCCCTAGGC
TGAGCCTGCTCATGTACCTCTGATTAATAAATGCTTATGAAATGA

<210> SEQ ID NO 134

<211> Length : 5,503

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 134

>R20779_T7

CTGTCCCCGGCCCCGGCGCGGGGGAGACGTGAGCGTGCACACGTACACACACAGCAGGGGAAGAGGCGCTCCAAGCG
GCGCCCAACTTTCTCCTTCCCTCCACGGGCCGGGTGAGAAAGTAGCCGGGGGCTATCCCGACCCGGCGGTTCTTGGG

GAGGGGGCCGAACAAGAAAAGGGAGGAGATGGAGATAACTTCCCCGGATTTAGCTTTTTTGTCTTTGTTTTTGTTCT
CACCACTTCCATCGGATGACTGGAGAGTAAAAGGGAACCCGGAGCGGGGTGGCGAGCAGCGCTTTGAGAAAATGCAG
GAGTGTGTTTGGAGACGCGTAAAGTTGCCTTTCAAGCTCTGGCCTCCGGGCACGCGATGCTCCGCGGCGGGCTGACT
CAGGGCTGCCTTGGGCCTCCCTGCCACCCTCCTGGAAATGATGCAAGTCCTGACTGTCACCTGGATCCCTGCAGCCC
AGCCTGGAATGCGTCTGGATTAGGGGAAAGACGAGAAACGACACTCCAGGTGTTGCACGGCCCACCAAAGCGGGAAG
ATAGGGCAGTTGCTCAGACCAAATACTGTATCTAGTGCTTCTGCTCCTATCTTCAATCGTGGGGTTCTTTTTAATGC
AAAGTGTCACAAGGCCAGGAATTCCCATGTGTGCTCAGTTGGCCCACAGCATCATTGTGCCTAGGAAACTGCTTCAA
TTTATCAAGTCCTCTGGGCTGGGAATCTCACTGAATTCCAAACGGCGGAAAGAGGAAACTTTCCCAACCCGATGTGG
GTGTGACGCGAGCCAGGGGCCCCAGGGACACTGTCCCAGAGCACACCGTCCCCCTTTAACAGCAACTGGAGCTTGGA
TTCGCTCTTATATTGTACAGTCCTTTCGACCATTGCCCTGGAGCACCCGCACACGCGCACGCATCTCCGGCCGCGCT
CACACACACTCATACACACGCACGCAAACGCGTGGCCGCCGCCAGGTCGGCAACTTTGTCCGGCGCTCCCAGCGGCG
CTCGGCTTCCTCCTGTAGTAGTTGAGCGCAGGCCCCGCCTCCCGGCCGTGTTGTCAAAAGGGCCGGGGTCTCGGATT
GGTCCAGCCGCCGGGACAACACCTGCTCGACTCCTTCATTCAAGTGACACCAGAGCTTCCAGGGATATTTGAGGCAC
CATCCCTGCCATTGCCGGGCACTCGCGGCGCTGCTAACGGCCTGGTCACATGCTCTCCGGAGAGCTACGGGAGGGCG
CTGGGTAACCTCTATCCGAGCCGCGGCCGCGAGGAGGAGGGAAAAGGCGAGCAAAAAGGAAGAGTGGGAGGAGGAGG
GGAAGCGGCGAAGGAGGAAGAGGAGGAGGAGGAAGAGGGGAGCACAAAGGATCCAGGTCTCCCGACGGGAGGTTAAT
ACCAAGAACCATGTGTGCCGAGCGGCTGGGCCAGTTCATGACCCTGGCTTTGGTGTTGGCCACCTTTGACCCGGCGC
GGGGGACCGACGCCACCAACCCACCCGAGGGTCCCCAAGACAGGAGCTCCCAGCAGAAAGGCCGCCTGTCCCTGCAG
AATACAGCGGAGATCCAGCACTGTTTGGTCAACGCTGGCGATGTGGGGTGTGGCGTGTTTGAATGTTTCGAGAACAA
CTCTTGTGAGATTCGGGGCTTACATGGGATTTGCATGACTTTTCTGCACAACGCTGGAAAATTTGATGCCCAGGGCA
AGTCATTCATCAAAGACGCCTTGAAATGTAAGGCCCACGCTCTGCGGCACAGGTTCGGCTGCATAAGCCGGAAGTGC
CCGGCCATCAGGGAAATGGTGTCCCAGTTGCAGCGGGAATGCTACCTCAAGCACGACCTGTGCGCGGCTGCCCAGGA
GAACACCCGGGTGATAGTGGAGATGATCCATTTCAAGGACTTGCTGCTGCACGAATGCTACAAGATAGAAATTACTA
TGCCCAAGAGGAGGAAAGTGAAGCTAAGAGATTAGAGAACTCGGACTGAGACCCTACGTGGACCTCGTGAACTTGCT
GCTGACCTGTGGGGAGGAGGTGAAGGAGGCCATCACCCACAGCGTGCAGGTTCAGTGTGAGCAGAACTGGGGAAGCC
TGTGCTCCATCTTGAGCTTCTGCACCTCGGCCATCCAGAAGCCTCCCACGGCGCCCCCCGAGCGCCAGCCCCAGGTG
GACAGAACCAAGCTCTCCAGGGCCCACCACGGGGAAGCAGGACATCACCTCCCAGAGCCCAGCAGTAGGGAGACTGG
CCGAGGTGCCAAGGGTGAGCGAGGTAGCAAGAGCCACCCAAACGCCCATGCCCGAGGCAGAGTCGGGGGCCTTGGGG
CTCAGGGACCTTCCGGAAGCAGCGAGTGGGAAGACGAACAGTCTGAGTATTCTGATATCCGGAGGTGAAATGAAAGG
CCTGGCCACGAAATCTTTCCTCCACGCCGTCCATTTTCTTATCTATGGACATTCCAAAACATTTACCATTAGAGAGG
GGGGATGTCACACGCAGGATTCTGTGGGGACTGTGGACTTCATCGAGGTGTGTGTTCGCGGAACGGACAGGTGAGAT
GGAGACCCCTGGGGCCGTGGGGTCTCAGGGGTGCCTGGTGAATTCTGCACTTACACGTACTCAAGGGAGCGCGCCCG
CGTTATCCTCGTACCTTTGTCTTCTTTCCATCTGTGGAGTCAGTGGGTGTCGGCCGCTCTGTTGTGGGGGAGGTGAA
CCAGGGAGGGGCAGGGCAAGGCAGGGCCCCCAGAGCTGGGCCACACAGTGGGTGCTGGGCCTCGCCCCGAAGCTTCT
GGTGCAGCAGCCTCTGGTGCTGTCTCCGCGGAAGTCAGGGCGGCTGGATTCCAGGACAGGAGTGAATGTAAAAATAA
ATATCGCTTAGAATGCAGGAGAAGGGTGGAGAGGAGGCAGGGGCCGAGGGGGTGCTTGGTGCCAAACTGAAATTCAG
TTTCTTGTGTGGGGCCTTGCGGTTCAGAGCTCTTGGCGAGGGTGGAGGGAGGAGTGTCATTTCTATGTGTAATTTCT
GAGCCATTGTACTGTCTGGGCTGGGGGGGACACTGTCCAAGGGAGTGGCCCCTATGAGTTTATATTTTAACCACTGC

TTCAAATCTCGATTTCACTTTTTTTATTTATCCAGTTATATCTACATATCTGTCATCTAAATAAATGGCTTTCAAAC
AAAGCAACTGGGTCATTAAAACCAGCTCAAAGGGGGTTTAAAAAAAAAAAAACCAGCCCATCCTTTGAGGCTGATTTT
TCTTTTTTTTAAGTTCTATTTTAAAAGCTATCAAACAGCGACATAGCCATACATCTGACTGCCTGACATGGACTCCT
GCCCACTTGGGGGAAACCTTATACCCAGAGGAAAATACACACCTGGGGAGTACATTTGACAAATTTCCCTTAGGATT
TCGTTATCTCACCTTGACCCTCAGCCAAGATTGGTAAAGCTGCGTCCTGGCGATTCCAGGAGACCCAGCTGGAAACC
TGGCTTCTCCATGTGAGGGGATGGGAAAGGAAAGAAGAGAATGAAGACTACTTAGTAATTCCCATCAGGAAATGCTG
ACCTTTTACATAAAATCAAGGAGACTGCTGAAAATCTCTAAGGGACAGGATTTTCCAGATCCTAATTGGAAATTTAG
CAATAAGGAGAGGAGTCCAAGGGGACAAATAAAGGCAGAGAGAAGAGACAGAACTAAAAATACGAGGAAAGGAGAGT
GAGGATTTTCATTAAAAGTCTCAGCAGTGGGTTTCTTGGGTTATTTAAAACATCACCTAAATAGGCCTTTTCTTCCT
AATTGGCCATCAAATTAAAGCCTATCCTTTCTCAAGCAGGAGCTGGTATTGTAGGGAGTGGCCGGGTATTCTGGGCT
GGGCTCTTCTGGAGTAGGGGGTCAGCAAACATTGTCTGCAAAGGGCCAGATACTGAATCCAGTACTTTCAGTTTGGC
GAGCCGTGAGGTCTCTGTCGAAACTACTCAACTCTGCCGTCCTAGCACAAAAGCAGCCATAGACAACACACAAACGA
GAGGGCTTGGCTCCCTTCCAGGAAGATTTATTTAACAGGCTCCCAGCTGAATTTCACTCACAGGACACAGTTTACTG
ATCTCTGTTCTAGTGAGTGGGTCAAAAAGCATATGCATCCTTATCCGTCAACTCATCAGCTCTTCCTCAAGGCAACC
TGAGGCCAGACACCAAGAAACCAAGCGTATCTGCTCTAAAATGACTTGTTCCTGGGGAATGCCTTCAACCAAAACAC
AGCTAGTATTTCTATGCCCCAAATCCAATCCCAGTCTTTCATGATCCATGCCGGCGGTTGGGTGGGGAGGGGAATCA
TTGGTTGGGGGAAGGGAGGAAACCCCACCTCCAGCCCCCGCCACCGGGCTCCCTGGGCACCCAGCAAGATCTGGGGC
TGCAGAGAACAGAAGAGCTGGTGCACTTAATCCAGCTCTGCCCTTGGGGGGAGGAGGACCTGTGTGTCAGGCTCTGC
CATGGGAACGAGTGTAAACCGTGGCTGTCTCCTGCAGTGAGCCACCGCGGCAGGCACGTTGACTGTTTTACTGACAT
CACTCAAAAGCTAAAGCAATAACATTCTCCTGCGTTGCTGAGTCAGCTGTTCATTTGTCCGCCAGCTCCTGGACTGG
ATGTGTGAAAGGCATCACATTTCCATTTTCCTCCGTGTAAATGTTTTATGTGTTCGCCTACTGATCCCATTCGTTGC
TTCTATTGTAAATATTTGTCATTTGTATTTATTATCTCTGTGTTTTCCCCCTAAGGCATAAAATGGTTTACTGTGTT
CATTTGAACCCATTTACTGATCTCTGTTGTATATTTTTCATGCCACTGCTTTGTTTTCTCCTCAGAAGTCGGGTAGA
TAGCATTTCTATCCCATCCCTCACGTTATTGGAAGCATGCAACAGTATTTATTGCTCAGGGTCTTCTGCTTAAAACT
GAGGAAGGTCCACATTCCTGCAAGCATTGATTGAGACATTTGCACAATCTAAAATGTAAGCAAAGTAGTCATTAAAA
ATACACCCTCTACTTGGGCTTTATACTGCATACAAATTTACTCATGAGCCTTCCTTTGAGGAAGGATGTGGATCTCC
AAATAAAGATTTAGTGTTTATTTTGAGCTCTGCATCTTAACAAGATGATCTGAACACCTCTCCTTTGTATCAATAAA
TAGCCCTGTTATTCTGAAGTGAGAGGACCAAGTATAGTAAAATGCTGACATCTAAAACTAAATAAATAGAAAACACC
AGGCCAGAACTATAGTCATACTCACACAAAGGGAGAAATTTAAACTCGAACCAAGCAAAAGGCTTCACGGAAATAGC
ATGGAAAAACAATGCTTCCAGTGGCCACTTCCTAAGGAGGAACAACCCCGTCTGATCTCAGAATTGGCACCACGTGA
GCTTGCTAAGTGATAATATCTGTTTCTACTACGGATTTAGGCAACAGGACCTGTACATTGTCACATTGCATTATTTT
TCTTCAAGCGTTAATAAAAGTTTTAAATAAATGGCT

<210> SEQ ID NO 135

<211> Length : 1,919

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 135

>R38144_PEA_2_T6

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGGGAGCGAGTCAAGGCCATGTTCTACCACGCCTAC
GACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGACCTCTCACCTGTGACGGGCACGACACCTGGGGCAG
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTG
AAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGTGGTAGGA
GGACTCCTGTCTGCTCATCTGCTCTCCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCCTCT
CCTGAGAATGGCTGAGGAGGCGGCCCGAAAACTCCTCCCAGCCTTTCAGACCCCCACTGGCATGCCATATGGAACAG
TGAACTTACTTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTTGAA
TTTGCCACCCTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTGGGA
GAGCCGGTCAGATATCGGGCTGGTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAGGCA
TCGGGGCTGGCGTGGACTCCTACTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATGGCC
ATGTTCCTAGAGTATAACAAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTACAA
GGGGACTGTGTCCATGCCAGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAGAGCCTCATTGGAGACATTG
ACAATGCCATGAGGACCTTCCTCAACTACTACACTGTATGGAAGCAGTTTGGGGGGCTCCCGGAATTCTACAACATT
CCTCAGGGATACACAGTGGAGAAGCGAGAGGGCTACCCACTTCGGCCAGAACTTATTGAAAGCGCAATGTACCTCTA
CCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGTGGAATCCATTGAAAAAATCAGCAAGGTGG
AGTGCGGATTTGCAACACTTGCTTCCTTCTCCCACATGTCAGATCAAAGATCTGCGAGACCACAAGCTGGACAACCG
CATGGAGTCGTTCTTCCTGGCCGAGACTGTGAAATACCTCTACCTCCTGTTTGACCCAACCAACTTCATCCACAACA
ATGGGTCCACCTTCGACGCGGTGATCACCCCCTATGGGGAGTGCATCCTGGGGGCTGGGGGGTACATCTTCAACACA
GAAGCTCACCCCATCGACCCTGCCGCCCTGCACTGCTGCCAGAGGCTGAAGGAAGAGCAGTGGGAGGTGGAGGACTT
GATGAGGGAATTCTACTCTCTCAAACGGAGCAGGTCGAAATTTCAGAAAAACACTGTTAGTTCGGGGCCATGGGAAC
CTCCAGCAAGGCCAGGAACACTCTTCTCACCAGAAAACCATGACCAGGCAAGGGAGAGGAAGCCTGCCAAACAGAAG
GTCCCACTTCTCAGCTGCCCCAGTCAGCCCTTCACCTCCAAGTTGGCATTACTGGGACAGGTTTTCCTAGACTCCTC
ATAACCACTGGATAATTTTTTTATTTTTATTTTTTTGAGGCTAAACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 136

<211> Length : 1,743

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 136

>R38144_PEA_2_T10

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGG
TTGAAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGTGGTA
GGAGGACTCCTGTCTGCTCATCTGCTCTCCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCC
TCTCCTGAGAATGGCTGAGGAGGCGGCCCGAAAACTCCTCCCAGCCTTTCAGACCCCCACTGGCATGCCATATGGAA
CAGTGAACTTACTTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTT
GAATTTGCCACCCTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTG
GGAGAGCCGGTCAGATATCGGGCTGGTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAG
GCATCGGGGCTGGCGTGGACTCCTACTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATG
GCCATGTTCCTAGAGTATAACAAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTA
CAAGGGGACTGTGTCCATGCCAGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAGAGCCTCATTGGAGACA
TTGACAATGCCATGAGGACCTTCCTCAACTACTACACTGTATGGAAGCAGTTTGGGGGGCTCCCGGAATTCTACAAC
ATTCCTCAGGGATACACAGTGGAGAAGCGAGAGGGCTACCCACTTCGGCCAGAACTTATTGAAAGCGCAATGTACCT
CTACCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGTGGAATCCATTGAAAAAATCAGCAAGG
TGGAGTGCGGATTTGCAACAATCAAAGATCTGCGAGACCACAAGCTGGACAACCGCATGGAGTCGTTCTTCCTGGCC
GAGACTGTGAAATACCTCTACCTCCTGTTTGACCCAACCAACTTCATCCACAACAATGGGTCCACCTTCGACGCGGT
GATCACCCCCTATGGGGAGTGCATCCTGGGGGCTGGGGGGTACATCTTCAACACAGAAGCTCACCCCATCGACCCTG
CCGCCCTGCACTGCTGCCAGAGGCTGAAGGAAGAGCAGTGGGAGGTGGAGGACTTGATGAGGGAATTCTACTCTCTC
AAACGGAGCAGGTCGAAATTTCAGAAAAACACTGTTAGTTCGGGGCCATGGGAACCTCCAGCAAGGCCAGGAACACT
CTTCTCACCAGAAAACCATGACCAGGCAAGGGAGAGGAAGCCTGCCAAACAGAAGGTCCCACTTCTCAGCTGCCCCA
GTCAGCCCTTCACCTCCAAGTTGGCATTACTGGGACAGGTTTTCCTAGACTCCTCATAACCACTGGATAATTTTTTT
ATTTTTATTTTTTTGAGGCTAAACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 137

<211> Length : 1,749

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 137

>R38144_PEA_2_T13

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGGGAGCGAGTCAAGGCCATGTTCTACCACGCCTAC

GACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGACCTCTCACCTGTGACGGGCACGACACCTGGGGCAG
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTG
AAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGTGGTAGGA
GGACTCCTGTCTGCTCATCTGCTCTCCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCCTCT
CCTGAGAATGGCTGAGGAGGCGGCCCGAAAACTCCTCCCAGCCTTTCAGACCCCCACTGGCATGCCATATGGAACAG
TGAACTTACTTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTTGAA
TTTGCCACCCTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTGGGA
GAGCCGGTCAGATATCGGGCTGGTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAGGCA
TCGGGGCTGGCGTGGACTCCTACTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATGGCC
ATGTTCCTAGAGTATAACAAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTACAA
GGGGACTGTGTCCATGCCAGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAGAACTTATTGAAAGCGCAAT
GTACCTCTACCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGTGGAATCCATTGAAAAAATCA
GCAAGGTGGAGTGCGGATTTGCAACAATCAAAGATCTGCGAGACCACAAGCTGGACAACCGCATGGAGTCGTTCTTC
CTGGCCGAGACTGTGAAATACCTCTACCTCCTGTTTGACCCAACCAACTTCATCCACAACAATGGGTCCACCTTCGA
CGCGGTGATCACCCCCTATGGGGAGTGCATCCTGGGGGCTGGGGGGTACATCTTCAACACAGAAGCTCACCCCATCG
ACCCTGCCGCCCTGCACTGCTGCCAGAGGCTGAAGGAAGAGCAGTGGGAGGTGGAGGACTTGATGAGGGAATTCTAC
TCTCTCAAACGGAGCAGGTCGAAATTTCAGAAAAACACTGTTAGTTCGGGGCCATGGGAACCTCCAGCAAGGCCAGG
AACACTCTTCTCACCAGAAAACCATGACCAGGCAAGGGAGAGGAAGCCTGCCAAACAGAAGGTCCCACTTCTCAGCT
GCCCCAGTCAGCCCTTCACCTCCAAGTTGGCATTACTGGGACAGGTTTTCCTAGACTCCTCATAACCACTGGATAAT
TTTTTTATTTTTATTTTTTTGAGGCTAAACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 138

<211> Length : 1,769

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 138

>R38144_PEA_2_T15

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGGGAGCGAGTCAAGGCCATGTTCTACCACGCCTAC
GACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGACCTCTCACCTGTGACGGGCACGACACCTGGGGCAG
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTG
AAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGTGGTAGGA
GGACTCCTGTCTGCTCATCTGCTCTCCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCCTCT
CCTGAGAATGGCTGAGGAGGCGGCCCGAAAACTCCTCCCAGCCTTTCAGACCCCCACTGGCATGCCATATGGAACAG

TGAACTTACTTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTTGAA
TTTGCCACCCTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTGGGA
GAGCCGGTCAGATATCGGGCTGGTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAGGCA
TCGGGGCTGGCGTGGACTCCTACTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATGGCC
ATGTTCCTAGAGCCTCATTGGAGACATTGACAATGCCATGAGGACCTTCCTCAACTACTACACTGTATGGAAGCAGT
TTGGGGGGCTCCCGGAATTCTACAACATTCCTCAGGGATACACAGTGGAGAAGCGAGAGGGCTACCCACTTCGGCCA
GAACTTATTGAAAGCGCAATGTACCTCTACCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGT
GGAATCCATTGAAAAAATCAGCAAGGTGGAGTGCGGATTTGCAACAATCAAAGATCTGCGAGACCACAAGCTGGACA
ACCGCATGGAGTCGTTCTTCCTGGCCGAGACTGTGAAATACCTCTACCTCCTGTTTGACCCAACCAACTTCATCCAC
AACAATGGGTCCACCTTCGACGCGGTGATCACCCCCTATGGGGAGTGCATCCTGGGGGCTGGGGGGTACATCTTCAA
CACAGAAGCTCACCCCATCGACCCTGCCGCCCTGCACTGCTGCCAGAGGCTGAAGGAAGAGCAGTGGGAGGTGGAGG
ACTTGATGAGGGAATTCTACTCTCTCAAACGGAGCAGGTCGAAATTTCAGAAAAACACTGTTAGTTCGGGGCCATGG
GAACCTCCAGCAAGGCCAGGAACACTCTTCTCACCAGAAAACCATGACCAGGCAAGGGAGAGGAAGCCTGCCAAACA
GAAGGTCCCACTTCTCAGCTGCCCCAGTCAGCCCTTCACCTCCAAGTTGGCATTACTGGGACAGGTTTTCCTAGACT
CCTCATAACCACTGGATAATTTTTTTATTTTTATTTTTTTGAGGCTAAACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 139

<211> Length : 1,522

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 139

>R38144_PEA_2_T19

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGGGAGCGAGTCAAGGCCATGTTCTACCACGCCTAC
GACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGACCTCTCACCTGTGACGGGCACGACACCTGGGGCAG
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTG
AAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGTGGTAGGA
GGACTCCTGTCTGCTCATCTGCTCTCCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCCTCT
CCTGAGAATGGCTGAGGAGGCGGCCCGAAAACTCCTCCCAGCCTTTCAGACCCCCACTGGCATGCCATATGGAACAG
TGAACTTACTTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTTGAA
TTTGCCACCCTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTGGGA
GAGCCGGTCAGATATCGGGCTGGTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAGGCA
TCGGGGCTGGCGTGGACTCCTACTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATGGCC
ATGTTCCTAGAGTATAACAAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTACAA

GGGGACTGTGTCCATGCCAGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAGAGCCTCATTGGAGACATTG
ACAATGCCATGAGGACCTTCCTCAACTACTACACTGTATGGAAGCAGTTTGGGGGGCTCCCGGAATTCTACAACATT
CCTCAGGGATACACAGTGGAGAAGCGAGAGGGCTACCCACTTCGGCCAGAACTTATTGAAAGCGCAATGTACCTCTA
CCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGTGGAATCCATTGAAAAAATCAGCAAGGTGG
AGTGCGGATTTGCAACAAAAAGATCTCGCTCTGTTGCCCAGGCTGGAGTGCAGTGGTGTGATCACGACTCACCGCAG
CCTTGACCTCCCACACTCAAGCAATCCTCCTGCCTTAGCCTTCCAAGTAGCTGGAACTCCAGGTGGTGGTTAATTTT
ATGTGTCAACCTGGCTGGACCACTGGGTACTCCGATATTTGGTCAAACATTATTCTGAG

<210> SEQ ID NO 140

<211> Length : 1,414

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 140

>R38144_PEA_2_T27

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTGCTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTG
CGCCAGGTCCCGACGGCTCCGCGCCAGATCCCGCCCACTACAGGGAGCGAGTCAAGGCCATGTTCTACCACGCCTAC
GACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGACCTCTCACCTGTGACGGGCACGACACCTGGGGCAG
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTGATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTG
AAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAACGCCTCTGTGTTTGAAACAAACATTCGAGAGTATAAC
AAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTACAAGGGGACTGTGTCCATGCC
AGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAGAGCCTCATTGGAGACATTGACAATGCCATGAGGACCT
TCCTCAACTACTACACTGTATGGAAGCAGTTTGGGGGGCTCCCGGAATTCTACAACATTCCTCAGGGATACACAGTG
GAGAAGCGAGAGGGCTACCCACTTCGGCCAGAACTTATTGAAAGCGCAATGTACCTCTACCGTGCCACGGGGGATCC
CACCCTCCTAGAACTCGGAAGAGATGCTGTGGAATCCATTGAAAAAATCAGCAAGGTGGAGTGCGGATTTGCAACAA
TCAAAGATCTGCGAGACCACAAGCTGGACAACCGCATGGAGTCGTTCTTCCTGGCCGAGACTGTGAAATACCTCTAC
CTCCTGTTTGACCCAACCAACTTCATCCACAACAATGGGTCCACCTTCGACGCGGTGATCACCCCCTATGGGGAGTG
CATCCTGGGGGCTGGGGGGTACATCTTCAACACAGAAGCTCACCCCATCGACCCTGCCGCCCTGCACTGCTGCCAGA
GGCTGAAGGAAGAGCAGTGGGAGGTGGAGGACTTGATGAGGGAATTCTACTCTCTCAAACGGAGCAGGTCGAAATTT
CAGAAAAACACTGTTAGTTCGGGGCCATGGGAACCTCCAGCAAGGCCAGGAACACTCTTCTCACCAGAAAACCATGA
CCAGGCAAGGGAGAGGAAGCCTGCCAAACAGAAGGTCCCACTTCTCAGCTGCCCCAGTCAGCCCTTCACCTCCAAGT
TGGCATTACTGGGACAGGTTTTCCTAGACTCCTCATAACCACTGGATAATTTTTTTATTTTTATTTTTTTGAGGCTA
AACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 141

<211> Length : 1,846

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 141

>HUMOSTRO_PEA_1_PEA_1_T14

GTGGCAGAAAACCTCATGACACAATCTCTCCGCCTCCCTGTGTTGGTGGAGGATGTCTGCAGCAGCATTTAAATTCT
GGGAGGGCTTGGTTGTCAGCAGCAGCAGGAGGAGGCAGAGCACAGCATCGTCGGGACCAGACTCGTCTCAGGCCAGT
TGCAGCCTTCTCAGCCAAACGCCGACCAAGGAAAACTCACTACCATGAGAATTGCAGTGATTTGCTTTTGCCTCCTA
GGCATCACCTGTGCCATACCAGTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAGCTTTACAACAAATACCC
AGATGCTGTGGCCACATGGCTAAACCCTGACCCATCTCAGAAGCAGAATCTCCTAGCCCCACAGGTATTTTTAAACT
TCTCATAATTAAACTACAGTGATGAAAGATAGCCACACTCAGGCCATTTGGGCTGCTCAGATGAATCCTGCCTGCCT
GCTGGCAAACATGTGCTTAGGACATTGACTGATCTGCCATGTTGGCTTCTCTCTGTGTTAAGCCATCCACAGATGAG
GCTGAAAAATAAAAACTGCTTTGGATTAAAAAGGTTAACTTTTGAATAAAAAAGCTAGGCATGTGTGATGCGCACTA
ACACGTGCCATTCCTTCTTCAGAATGCTGTGTCCTCTGAAGAAACCAATGACTTTAAACAAGAGACCCTTCCAAGTA
AGTCCAACGAAAGCCATGACCACATGGATGATATGGATGATGAAGATGATGATGACCATGTGGACAGCCAGGACTCC
ATTGACTCGAACGACTCTGATGATGTAGATGACACTGATGATTCTCACCAGTCTGATGAGTCTCACCATTCTGATGA
ATCTGATGAACTGGTCACTGATTTTCCCACGGACCTGCCAGCAACCGAAGTTTTCACTCCAGTTGTCCCCACAGTAG
ACACATATGATGGCCGAGGTGATAGTGTGGTTTATGGACTGAGGTCAAAATCTAAGAAGTTTCGCAGACCTGACATC
CAGTACCCTGATGCTACAGACGAGGACATCACCTCACACATGGAAAGCGAGGAGTTGAATGGTGCATACAAGGCCAT
CCCCGTTGCCCAGGACCTGAACGCGCCTTCTGATTGGGACAGCCGTGGGAAGGACAGTTATGAAACGAGTCAGCTGG
ATGACCAGAGTGCTGAAACCCACAGCCACAAGCAGTCCAGATTATATAAGCGGAAAGCCAATGATGAGAGCAATGAG
CATTCCGATGTGATTGATAGTCAGGAACTTTCCAAAGTCAGCCGTGAATTCCACAGCCATGAATTTCACAGCCATGA
AGATATGCTGGTTGTAGACCCCAAAAGTAAGGAAGAAGATAAACACCTGAAATTTCGTATTTCTCATGAATTAGATA
GTGCATCTTCTGAGGTCAATTAAAAGGAGAAAAAATACAATTTCTCACTTTGCATTTAGTCAAAAGAAAAAATGCTT
TATAGCAAAATGAAAGAGAACATGAAATGCTTCTTTCTCAGTTTATTGGTTGAATGTGTATCTATTTGAGTCTGGAA
ATAACTAATGTGTTTGATAATTAGTTTAGTTTGTGGCTTCATGGAAACTCCCTGTAAACTAAAAGCTTCAGGGTTAT
GTCTATGTTCATTCTATAGAAGAAATGCAAACTATCACTGTATTTTAATATTTGTTATTCTCTCATGAATAGAAATT
TATGTAGAAGCAAACAAAATACTTTTACCCACTTAAAAAGAGAATATAACATTTTATGTCACTATAATCTTTTGTTT
TTTAAGTTAGTGTATATTTTGTTGTGATTATCTTTTTGTGGTGTGAATAAATCTTTTATCTTGAATGTAATAAGA

<210> SEQ ID NO 142

<211> Length : 1,769

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 142

>HUMOSTRO_PEA_1_PEA_1_T16

GTGGCAGAAAACCTCATGACACAATCTCTCCGCCTCCCTGTGTTGGTGGAGGATGTCTGCAGCAGCATTTAAATTCT
GGGAGGGCTTGGTTGTCAGCAGCAGCAGGAGGAGGCAGAGCACAGCATCGTCGGGACCAGACTCGTCTCAGGCCAGT
TGCAGCCTTCTCAGCCAAACGCCGACCAAGGAAAACTCACTACCATGAGAATTGCAGTGATTTGCTTTTGCCTCCTA
GGCATCACCTGTGCCATACCAGTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAGCACTAAAGATGTACCTA
CCCCTCCACAACAGATGAAACTGTGCCAGCCAAACAACAAATGGGCATTGTCCCCAGAAGCTTGGACAAAAAGGCAC
ACAGAGTTCAATTCCAGTTGAACAGAATAAAGGCCAAAATAGAGCTGCCTTGGGGGTCACTGCAATTAGACTGCTTA
ATGAAGACATTAAAAGAACTTTACAACAAATACCCAGATGCTGTGGCCACATGGCTAAACCCTGACCCATCTCAGAA
GCAGAATCTCCTAGCCCCACAGAATGCTGTGTCCTCTGAAGAAACCAATGACTTTAAACAAGAGACCCTTCCAAGTA
AGTCCAACGAAAGCCATGACCACATGGATGATATGGATGATGAAGATGATGATGACCATGTGGACAGCCAGGACTCC
ATTGACTCGAACGACTCTGATGATGTAGATGACACTGATGATTCTCACCAGTCTGATGAGTCTCACCATTCTGATGA
ATCTGATGAACTGGTCACTGATTTTCCCACGGACCTGCCAGCAACCGAAGTTTTCACTCCAGTTGTCCCCACAGTAG
ACACATATGATGGCCGAGGTGATAGTGTGGTTTATGGACTGAGGTCAAAATCTAAGAAGTTTCGCAGACCTGACATC
CAGTACCCTGATGCTACAGACGAGGACATCACCTCACACATGGAAAGCGAGGAGTTGAATGGTGCATACAAGGCCAT
CCCCGTTGCCCAGGACCTGAACGCGCCTTCTGATTGGGACAGCCGTGGGAAGGACAGTTATGAAACGAGTCAGCTGG
ATGACCAGAGTGCTGAAACCCACAGCCACAAGCAGTCCAGATTATATAAGCGGAAAGCCAATGATGAGAGCAATGAG
CATTCCGATGTGATTGATAGTCAGGAACTTTCCAAAGTCAGCCGTGAATTCCACAGCCATGAATTTCACAGCCATGA
AGATATGCTGGTTGTAGACCCCAAAAGTAAGGAAGAAGATAAACACCTGAAATTTCGTATTTCTCATGAATTAGATA
GTGCATCTTCTGAGGTCAATTAAAAGGAGAAAAAATACAATTTCTCACTTTGCATTTAGTCAAAAGAAAAAATGCTT
TATAGCAAAATGAAAGAGAACATGAAATGCTTCTTTCTCAGTTTATTGGTTGAATGTGTATCTATTTGAGTCTGGAA
ATAACTAATGTGTTTGATAATTAGTTTAGTTTGTGGCTTCATGGAAACTCCCTGTAAACTAAAAGCTTCAGGGTTAT
GTCTATGTTCATTCTATAGAAGAAATGCAAACTATCACTGTATTTTAATATTTGTTATTCTCTCATGAATAGAAATT
TATGTAGAAGCAAACAAAATACTTTTACCCACTTAAAAAGAGAATATAACATTTTATGTCACTATAATCTTTTGTTT
TTTAAGTTAGTGTATATTTTGTTGTGATTATCTTTTTGTGGTGTGAATAAATCTTTTATCTTGAATGTAATAAGA

<210> SEQ ID NO 143

<211> Length : 378

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 143

>HUMOSTRO_PEA_1_PEA_1_T30

GTGGCAGAAAACCTCATGACACAATCTCTCCGCCTCCCTGTGTTGGTGGAGGATGTCTGCAGCAGCATTTAAATTCT
GGGAGGGCTTGGTTGTCAGCAGCAGCAGGAGGAGGCAGAGCACAGCATCGTCGGGACCAGACTCGTCTCAGGCCAGT
TGCAGCCTTCTCAGCCAAACGCCGACCAAGGAAAACTCACTACCATGAGAATTGCAGTGATTTGCTTTTGCCTCCTA
GGCATCACCTGTGCCATACCAGTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAGGTAAGCATCTTTTATGT
TTTTATATAGTTAAATCATTTACTCAATTATGGCGAGAGGTGCAAGAAACGTATTTGCTGCGATATTACT

<210> SEQ ID NO 144

<211> Length : 1,295

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 144

>R11723_PEA_1_T15

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC
GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACG
ACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGATGGAGCAA
AGTGCCGGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGGCCTGTCTCATCGCCTCTGCCGGTTCTCCTTGCAG
AGGACTGGCGCCGGGACGCGAAGAGCAACGGGCGCTGCACAAAGCGGGCGCTGTCGGTGGTGGAGTGCGCATGTACG
CGCAGGCGCTTCTCGTGGTTGGCGTGCTGCAGCGACAGGCGGCAGCACAGCACCTGCACGAACACCCGCCGAAACTG
CTGCGAGGACACCGTGTACAGGAGCGGGTTGATGACCGAGCTGAGGTAGAAAAACGTCTCCGAGAAGGGGAGGAGGA
TCATGTACGCCCGGAAGTAGGACCTCGTCCAGTCGTGCTTGGGTTTGGCCGCAGCCATGATCCTCCGAATCTGGTTG
GGCATCCAGCATACGGCCAATGTCACAACAATCAGCCCTGGGCAGACACGAGCAGGAGGGAGAGACAGAGAAAAGAA
AAACACAGCATGAGAACACAGTAAATAAATAAAACCATAAAATATTTAGCCCCTCTGTTCTGTGCTTACTGGCCAGG
AAATGGTACCAATTTTTCAGTGTTGGACTTGACAGCTTCTTTTGCCACAAGCAAGAGAGAATTTAACACTGTTTCAA
ACCCGGGGGAGTTGGCTGTGTTAAAGAAAGACCATTAAATGCTTTAGACAGTGTATTTATACC

<210> SEQ ID NO 145

<211> Length : 1,367

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 145

>R11723_PEA_1_T17

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC
GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACG
ACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGATGGAGCAA
AGTGCCGGGTCTCACTGTGTCACGAGGCTGGAGTGCAGTGGAACAATTTCAGCACACTGCAACCTCTGCCTCCCAGG
CTCAAATGATCATCCCACCTAAGCCTCCGGAGTAGCTGGGACCACAGGCAAGCGCCACCATGCCCAGCTGATACCAA
TGTCTTTTAAAAAATGTTGTATGTGGAAATAAATTGAGACTTATAGAAAAGCTGCAAAAATAGTGCAGTTTCTATAT
ATCCTTCCCCCATCTTTGGCTAGTGTTAACAATCTACATAACCGCAGTACGATGATCAAGGCTAGGAAATTAACATT
GGCACAGTACTGTTAATGAAACCATGCTTTGTTTTGAGATTCCCACAGTTTTGCCTTTTTCTGTTCCAAGATCCTAT
CCAGGATCCCACGTTGCATTTCATTGTCATGTCTCCTTCTCCTCTAACCTCTGACAATGCATCATTCTTTCCATGTC
TTTTGTGATGTTGACACTTTTGAAGAGGACTGGTCCAGATTTTTGTACACTGTCCCTCAGTTTGGGATTGTCTGCTG
TTTTCTCATGAACAGATAGAGGTTTTGCATTTTTGACAAGAATCCTCAGAAGAGATGCACCCTTCTCAGTGCACTGT
AGCAAGGGGCGCATGCTGTCAATGTCTTACTGGTGATGTTAACTTTGATCGCTTTTGATTCAGATAGTATCTGCTGG
GTTTTTCCACTGTAAAGTTACTATTTTTTCCATTGTAATTAATAAATAACTTGAGGGA

<210> SEQ ID NO 146

<211> Length : 1,520

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 146

>R11723_PEA_1_T19

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC

GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACG
ACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGATGGAGCAA
AGTGCCGACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCAGTTGACCACTTAGTTCTC
AAGAAGCAACTATCTCTTTCATGTGCCTTCTGAGGAAGTATTCAGAGGGGGAATATCAAATGTCTTTCCCTTGGACT
CTCCCAGGTCTCACTGTGTCACGAGGCTGGAGTGCAGTGGAACAATTTCAGCACACTGCAACCTCTGCCTCCCAGGC
TCAAATGATCATCCCACCTAAGCCTCCGGAGTAGCTGGGACCACAGGCAAGCGCCACCATGCCCAGCTGATACCAAT
GTCTTTTAAAAAATGTTGTATGTGGAAATAAATTGAGACTTATAGAAAAGCTGCAAAAATAGTGCAGTTTCTATATA
TCCTTCCCCCATCTTTGGCTAGTGTTAACAATCTACATAACCGCAGTACGATGATCAAGGCTAGGAAATTAACATTG
GCACAGTACTGTTAATGAAACCATGCTTTGTTTTGAGATTCCCACAGTTTTGCCTTTTTCTGTTCCAAGATCCTATC
CAGGATCCCACGTTGCATTTCATTGTCATGTCTCCTTCTCCTCTAACCTCTGACAATGCATCATTCTTTCCATGTCT
TTTGTGATGTTGACACTTTTGAAGAGGACTGGTCCAGATTTTTGTACACTGTCCCTCAGTTTGGGATTGTCTGCTGT
TTTCTCATGAACAGATAGAGGTTTTGCATTTTTGACAAGAATCCTCAGAAGAGATGCACCCTTCTCAGTGCACTGTA
GCAAGGGGCGCATGCTGTCAATGTCTTACTGGTGATGTTAACTTTGATCGCTTTTGATTCAGATAGTATCTGCTGGG
TTTTTCCACTGTAAAGTTACTATTTTTTCCATTGTAATTAATAAATAACTTGAGGGA

<210> SEQ ID NO 147

<211> Length : 1,371

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 147

>R11723_PEA_1_T20

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC
GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACG
ACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGATGGAGCAA
AGTGCCGACAGGGTCTCACTGTGTCACGAGGCTGGAGTGCAGTGGAACAATTTCAGCACACTGCAACCTCTGCCTCC
CAGGCTCAAATGATCATCCCACCTAAGCCTCCGGAGTAGCTGGGACCACAGGCAAGCGCCACCATGCCCAGCTGATA
CCAATGTCTTTTAAAAAATGTTGTATGTGGAAATAAATTGAGACTTATAGAAAAGCTGCAAAAATAGTGCAGTTTCT
ATATATCCTTCCCCCATCTTTGGCTAGTGTTAACAATCTACATAACCGCAGTACGATGATCAAGGCTAGGAAATTAA
CATTGGCACAGTACTGTTAATGAAACCATGCTTTGTTTTGAGATTCCCACAGTTTTGCCTTTTTCTGTTCCAAGATC

CTATCCAGGATCCCACGTTGCATTTCATTGTCATGTCTCCTTCTCCTCTAACCTCTGACAATGCATCATTCTTTCCA
TGTCTTTTGTGATGTTGACACTTTTGAAGAGGACTGGTCCAGATTTTTGTACACTGTCCCTCAGTTTGGGATTGTCT
GCTGTTTTCTCATGAACAGATAGAGGTTTTGCATTTTTGACAAGAATCCTCAGAAGAGATGCACCCTTCTCAGTGCA
CTGTAGCAAGGGGCGCATGCTGTCAATGTCTTACTGGTGATGTTAACTTTGATCGCTTTTGATTCAGATAGTATCTG
CTGGGTTTTTCCACTGTAAAGTTACTATTTTTTCCATTGTAATTAATAAATAACTTGAGGGA

<210> SEQ ID NO 148

<211> Length : 2,213

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 148

>R11723_PEA_1_T5

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC
GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAGCTGAACAACG
ACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGTGATGGAGCAA
AGTGCCGACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCAGTTGACCACTTAGTTCTC
AAGAAGCAACTATCTCTTTCATGTGCCTTCTGAGGAAGTATTCAGAGGGGGAATATCAAATGTCTTTCCCTTGGACT
CTCCCAGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGGCCTGTCTCATCGCCTCTGCCGGGTACCAGTCCTTC
TGCTCCCCAGGGAAACTGAACTCAGTTTGCATCAGCTGCTGCAACACCCCTCTTTGTAACGGGCCAAGGCCCAAGAA
AAGGGGAAGTTCTGCCTCGGCCCTCAGGCCAGGGCTCCGCACCACCATCCTGTTCCTCAAATTAGCCCTCTTCTCGG
CACACTGCTGAAGCTGAAGGAGATGCCACCCCCTCCTGCATTGTTCTTCCAGCCCTCGCCCCCAACCCCCCACCTCC
CTGAGTGAGTTTCTTCTGGGTGTCCTTTTATTCTGGGTAGGGAGCGGGAGTCCGTGTTCTCTTTTGTTCCTGTGCAA
ATAATGAAAGAGCTCGGTAAAGCATTCTGAATAAATTCAGCCTGACTGAATTTTCAGTATGTACTTGAAGGAAGGAG
GTGGAGTGAAAGTTCACCCCCATGTCTGTGTAACCGGAGTCAAGGCCAGGCTGGCAGAGTCAGTCCTTAGAAGTCAC
TGAGGTGGGCATCTGCCTTTTGTAAAGCCTCCAGTGTCCATTCCATCCCTGATGGGGGCATAGTTTGAGACTGCAGA
GTGAGAGTGACGTTTTCTTAGGGCTGGAGGGCCAGTTCCCACTCAAGGCTCCCTCGCTTGACATTCAAACTTCATGC
TCCTGAAAACCATTCTCTGCAGCAGAATTGGCTGGTTTCGCGCCTGAGTTGGGCTCTAGTGACTCGAGACTCAATGA
CTGGGACTTAGACTGGGGCTCGGCCTCGCTCTGAAAAGTGCTTAAGAAAATCTTCTCAGTTCTCCTTGCAGAGGACT
GGCGCCGGGACGCGAAGAGCAACGGGCGCTGCACAAAGCGGGCGCTGTCGGTGGTGGAGTGCGCATGTACGCGCAGG
CGCTTCTCGTGGTTGGCGTGCTGCAGCGACAGGCGGCAGCACAGCACCTGCACGAACACCCGCCGAAACTGCTGCGA

GGACACCGTGTACAGGAGCGGGTTGATGACCGAGCTGAGGTAGAAAAACGTCTCCGAGAAGGGGAGGAGGATCATGT
ACGCCCGGAAGTAGGACCTCGTCCAGTCGTGCTTGGGTTTGGCCGCAGCCATGATCCTCCGAATCTGGTTGGGCATC
CAGCATACGGCCAATGTCACAACAATCAGCCCTGGGCAGACACGAGCAGGAGGGAGAGACAGAGAAAAGAAAAACAC
AGCATGAGAACACAGTAAATAAATAAAACCATAAAATATTTAGCCCCTCTGTTCTGTGCTTACTGGCCAGGAAATGG
TACCAATTTTTCAGTGTTGGACTTGACAGCTTCTTTTGCCACAAGCAAGAGAGAATTTAACACTGTTTCAAACCCGG
GGGAGTTGGCTGTGTTAAAGAAAGACCATTAAATGCTTTAGACAGTGTATTTATACC

<210> SEQ ID NO 149

<211> Length : 2,247

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 149

>R11723_PEA_1_T6

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
TCTCCGCGGCAGCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAGCCTCCGCTGCTGTC
GCCTCCTCTGATGCGCTTGCCCTCTCCCGGCCCCGGGACTCCGGGAGAATGTGGGTCCTAGGCATCGCGGCAACTTT
TTGCGGATTGTTCTTGCTTCCAGGTGAGAATACCCAGAGGCCAGCAGCCGAGGCCAGGCTTTGCGCTGCAAATCCAG
TGCTACCAGTGTGAAGAATTCCAGCTGAACAACGACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGT
TCAAGACATGTGTCAGAAAGAAGTGATGGAGCAAAGTGCCGACACTAAAAGAACAAACACCTTGCTCTTCGAGATGA
GACATTTTGCCAAGCAGTTGACCACTTAGTTCTCAAGAAGCAACTATCTCTTTCATGTGCCTTCTGAGGAAGTATTC
AGAGGGGGAATATCAAATGTCTTTCCCTTGGACTCTCCCAGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGGC
CTGTCTCATCGCCTCTGCCGGGTACCAGTCCTTCTGCTCCCCAGGGAAACTGAACTCAGTTTGCATCAGCTGCTGCA
ACACCCCTCTTTGTAACGGGCCAAGGCCCAAGAAAAGGGGAAGTTCTGCCTCGGCCCTCAGGCCAGGGCTCCGCACC
ACCATCCTGTTCCTCAAATTAGCCCTCTTCTCGGCACACTGCTGAAGCTGAAGGAGATGCCACCCCCTCCTGCATTG
TTCTTCCAGCCCTCGCCCCCAACCCCCCACCTCCCTGAGTGAGTTTCTTCTGGGTGTCCTTTTATTCTGGGTAGGGA
GCGGGAGTCCGTGTTCTCTTTTGTTCCTGTGCAAATAATGAAAGAGCTCGGTAAAGCATTCTGAATAAATTCAGCCT
GACTGAATTTTCAGTATGTACTTGAAGGAAGGAGGTGGAGTGAAAGTTCACCCCCATGTCTGTGTAACCGGAGTCAA
GGCCAGGCTGGCAGAGTCAGTCCTTAGAAGTCACTGAGGTGGGCATCTGCCTTTTGTAAAGCCTCCAGTGTCCATTC
CATCCCTGATGGGGGCATAGTTTGAGACTGCAGAGTGAGAGTGACGTTTTCTTAGGGCTGGAGGGCCAGTTCCCACT
CAAGGCTCCCTCGCTTGACATTCAAACTTCATGCTCCTGAAAACCATTCTCTGCAGCAGAATTGGCTGGTTTCGCGC
CTGAGTTGGGCTCTAGTGACTCGAGACTCAATGACTGGGACTTAGACTGGGGCTCGGCCTCGCTCTGAAAAGTGCTT
AAGAAAATCTTCTCAGTTCTCCTTGCAGAGGACTGGCGCCGGGACGCGAAGAGCAACGGGCGCTGCACAAAGCGGGC

GCTGTCGGTGGTGGAGTGCGCATGTACGCGCAGGCGCTTCTCGTGGTTGGCGTGCTGCAGCGACAGGCGGCAGCACA
GCACCTGCACGAACACCCGCCGAAACTGCTGCGAGGACACCGTGTACAGGAGCGGGTTGATGACCGAGCTGAGGTAG
AAAAACGTCTCCGAGAAGGGGAGGAGGATCATGTACGCCCGGAAGTAGGACCTCGTCCAGTCGTGCTTGGGTTTGGC
CGCAGCCATGATCCTCCGAATCTGGTTGGGCATCCAGCATACGGCCAATGTCACAACAATCAGCCCTGGGCAGACAC
GAGCAGGAGGGAGAGACAGAGAAAAGAAAAACACAGCATGAGAACACAGTAAATAAATAAAACCATAAAATATTTAG
CCCCTCTGTTCTGTGCTTACTGGCCAGGAAATGGTACCAATTTTTCAGTGTTGGACTTGACAGCTTCTTTTGCCACA
AGCAAGAGAGAATTTAACACTGTTTCAAACCCGGGGGAGTTGGCTGTGTTAAAGAAAGACCATTAAATGCTTTAGAC
AGTGTATTTATACC

<210> SEQ ID NO 150

<211> Length : 876

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 150

>R16276_PEA_1_T6

GTGGAGGAGGATGGTGGGGAGTGGTGGTGTCTTCGTCCTGGGAGAAGGCGAAGCAACTTCCAGGAGGAAACGGGCGT
TTCCTTCCCACGCGCTCGAGCGAGCCCTGGGTCCTGGCCTCGGAACTCCACCCAGCCCCTCCCCACCCTCTGGGAAA
AGCCAGTCGCCACACACAGGCACACGCAGGCCCCGGCGCCGCGCCCTAAGGAGAGCAGCACCCACAGCCAATTGCCA
TGGCAACCCCGGGGTTCGTTCCACTTCCCCACCCAGCCGATCTCCCCCCTCCTCCCTGCACTGCAGCCAACCGGCTT
GTGCGCGTCCCAGGAGCGCGCTATAAAACCTGTGCTGGGCGTGATCGGCAAGCACCGGACCAGGGGGAAGGCGAGCA
GTGCCAATCTACAGCGAAGAAAGTCTCGTTTGGTAAAAGCGAGAAGGGAAAGCCTGAGCATGCAGAGTGTGCAGAGC
ACGAGCTTTTGTCTCCGAAAGCAGTGCCTTTGCCTGACCTTCCTGCTTCTCCATCTCCTGGGACAGGTCGCTGCGAC
TCAGCGCTGCCCTCCCCAGTGCCCGGGCCAGTGCCCTGCGACGCCGCCGACCTGCGCCCCCGGGGTGCGCGCGGTGC
TGGACGGCTGCTCATGCTGTCTGGTGTGTGCCCGCCAGCGTGGCGAGAGCTGCTCAGATCTGGAGCCATGCGACGAG
AGCAGTGGCCTCTACTGTGATCGCAGCGCGGACCCCAGCAACCAGACTGGCATCTGCACGGGTAATCCTGCTCCCTC
TGCTGTTTGACCTCTTCTCCTGCAGCTAAGTGAAGCTGCTTCCTCCCTTCTCTTTTGTATTCCCCTTCCCAGAGGGC
GATAAGCAAATAATAATAATGCAATAAAT

Segment nucleic acid sequences

<210> SEQ ID NO 151

<211> Length : 232

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 151
>H61775_node_2
ATCTGGTGGTTCTCCGGAGAGCAGCTTCCTTGGGTGTTACATGAGCCAAGCCCTCACTGTACAGAAGAGTGAGAGCT
GAAACCTGTTCCCTGAGCTGATCAGAAGGACATCCCTTGGCCCCTCCATCTGGGCTCCTGTGGATAGGAGGGGCTGG
GTGAGCAGGCCAGCTGGGCTATGGTGTGGTGCCTCGGCCTGGCCGTCCTCAGCCTGGTCATCAGCCAGGGGGCTGAC
G

<210> SEQ ID NO 152

<211> Length : 189

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 189
>H61775_node_4
GTCGAGGGAAGCCTGAGGTGGTATCGGTGGTGGGCCGGGCTGGGGAGAGTGTGGTGCTGGGCTGTGACCTGCTGCCC
CCGGCCGGCCGGCCCCCCCCTGCATGTCATCGAGTGGCTGCGCTTTGGATTCCTGCTTCCCATCTTCATCCAGTTCGG
CCTCTACTCTCCCCGAATTGACCCTGATTACGTGG

<210> SEQ ID NO 153

<211> Length : 201

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 153
>H61775_node_6

GTTCTCCTCAGGTGGGAGGTAGGGAGGTATCAGCAAGAAAGGTGGGCTGGGTAGAGTCGCACAAGGCCTCCTATGAA
CGGCTTTGTCCCTGCTCTGATCTCATCTCCAGCTCTGCTGCCTTAACTCTGCTTAATAAGCATGGCTGTGCTCCCAA
GCAGTGTTAATTCATTGAAAGATGTCATTCATTTACACACACACACA

<210> SEQ ID NO 154

<211> Length : 698

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 154
>H61775_node_8
GTGACTGTGGGTTCCCTGCCTTCCGAGAGCTTAAGAGAGCAGAGACTGTGTCTCCTGTTTTCTTCACACGCCGCTGC
ATATGGGAAGATCTGAAGTCAACAGGCTTTAGCCCTGCAGGTGGAGGGAGGCCTCCAGGAGGTGGGCCCAGGACTCA
GGAGGACTCAGGGCTGCCCTGCTGGCGATCTTCCTGTTCTGTAACACTACAGGTCTAGCAGTCCAGCTGTCACAGAA
AAGCTAGGACATGCAGTATGCTTCTTTGGATATTCTGAGTAACATTTGGACTGTTACCCATTGGCTACCAGCATCTC
CCAAGTGAGAATACATAGATTACCCCCAGTGCCCTGAACAGCACTCGGTCCTAACACCCGTGTCCATGGAAAGCACG
CCGCGTCTGGAGAAAGAAGCCGAAGGCTCTTGTCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATT
CCTTCAGCTACAATGGAATTCTTGGGAGGATTAAATATGGTGACAACGCCTAATATTAGATGGCCTGTATTCCACAC
TCAATCTTCCTTCCCTCTTCTTCCTTCTTTGTAGAGCTATAATGAAAAGTATCATGTGGGACACAGAAGAGGTTGCA
GTCTGGGGTCTGCAGGGCTTAGCGGCCAGGCAGATTAGCTTTCTTGAGGAATCCTGACAGTGGGTGGAAGGGTATGA
TGATG

<210> SEQ ID NO 155

<211> Length : 86

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 155
>H61775_node_0
GGAGGCGCTCGGGGCATCCGAGGCGGGGAGGCGGGTCCGCCCCCTATTGTGTAGCGGCGAGAGTGGAGCCGAGCGGT
GCGGAGCAG

<210> SEQ ID NO 156

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 156
>H61775_node_5
GATAAGA

<210> SEQ ID NO 157

<211> Length : 203

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 157
>M85491_PEA_1_node_0
TCTGCTGGCTGCGCGGTGGCGGCGGCTGTGTGTGCGCCGCGCCTTGCCGCCCCCCCTGGCCCCCCGAGCCCGGGGCG
CGCGCTCCCGCCCGGGCCGTCCGGGCCCCGCGGCGCCGCGGCCCGAGGCCCCGGGAAGCGCAGCCATGGCTCTGCGG
AGGCTGGGGGCCGCGCTGCTGCTGCTGCCGCTGCTCGCCGCCGTGGAAG

<210> SEQ ID NO 158

<211> Length : 229

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 158
>M85491_PEA_1_node_13
AGAGACAGGATCTCACTATGTTGTCCAGGCTGGTCTTGAACTCGTGGCCACAAATGATCCTCCCACCTCAGCCTCCC
AAAGTGTTGGAATTATAGGCATGAACCACCATGCCCAGGAGGAGAATTTTTGATAATAATATTTTGTGGACATCTTT
GCATATCATGTCAGAGCTATAACATCATTGTGGAGAAGCTCTTAGGATCCCATAGAATAAATGTACCGTAATTTA

<210> SEQ ID NO 159

<211> Length : 336

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 159

>M85491_PEA_1_node_21

CCATCCCCTCCGCGCCCCAGGCTGTGATTTCCAGTGTCAATGAGACCTCCCTCATGCTGGAGTGGACCCCTCCCCGC
GACTCCGGAGGCCGAGAGGACCTCGTCTACAACATCATCTGCAAGAGCTGTGGCTCGGGCCGGGGTGCCTGCACCCG
CTGCGGGGACAATGTACAGTACGCACCACGCCAGCTAGGCCTGACCGAGCCACGCATTTACATCAGTGACCTGCTGG
CCCACACCCAGTACACCTTCGAGATCCAGGCTGTGAACGGCGTTACTGACCAGAGCCCCTTCTCGCCTCAGTTCGCC
TCTGTGAACATCACCACCAACCAGGCAG

<210> SEQ ID NO 160

<211> Length : 125

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 160

>M85491_PEA_1_node_23

CTCCATCGGCAGTGTCCATCATGCATCAGGTGAGCCGCACCGTGGACAGCATTACCCTGTCGTGGTCCCAGCCAGAC
CAGCCCAATGGCGTGATCCTGGACTATGAGCTGCAGTACTATGAGAAG

<210> SEQ ID NO 161

<211> Length : 1,305

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 161

>M85491_PEA_1_node_24

GTACCTATTGGCTGGGTGCTGTCCCCATCACCCACCTCCCTGAGGGCCCCTCTCCCAGGCTGAGGCCTGGGAGTTCT
GCCCCACCGCAAGATGAGACGCACTGGTGCAGCAGAAAGAGCACTGGCCTTGGAGTCAGGCTGCCTGGCTCCCAATC
CAGCTCCGCTCCTTCCCACTGTGAGACCTCAGGCAGGTGCCTTGACCTCTCTGGATCTCACTTTTCTGGTCTGGAGG

ATACACCCAGCAATCTCAGTGAAATGCAACAGTCACATCCCTTTCCCTACCACGACCCTTTCATCTTGACCTCAGTG
GCTTGATGTTGGGAAAAACTGGGTTTCCAAAAAGCTGCACTTATGAAGTGATAATTAGTCACTCACCTCTTCTTCGA
CAGAGATTTGAAACAGCTCAAGAGAGCTTCCGCCTGCCCTGCTCTGAGTCCTGCTAAAACACCCACTTTCACTCGCC
TGCATGCCCTTTGCATGGGGAGAGGTGATTTCACTTTGAGCTTTTAAATCAGACCTTAATTACTCCCTTTGGGTGGA
AGCCCCTGGGATGGTAGAAGGATCACTGGACTAAGAGTGAGAAGCCGTAGGTTCAAATCCCAGCTCCGTCCTTCACC
AGCTATGTGACCTTGGGCAGGCGTCTTTCTCCCTCTGAACCTCAGTTTCCACCTGTGTCGAGTGTGGGTGAGACCCC
TCGCGGGGAGCTATGCAGGTTACGGAGAAAAGGCAGCACAGCACCCAGAATGGGACCTGGCCCTCAGCAGAGGCCAT
GTGTGTCCCTGGCCTTCCTCCTCTGCCCTGCCTGCTGCACAGTGGGCAATGGTGACAGGATGGGAGGCCAAGTGGAT
GTGGGGTCTGCACAGTACAGGGGCCAGGAGGTAGACAGCACAATTGCCCACCCACATGGCTGGACATCAGAGGCCCC
AGGAAGCCTCTCCTTTGAATGATCACTTCTCTTACCTGCTCCAGGAGGCAACAAACAGCCACAGAGGCTGCAAGGGC
ACCTGGGAAAGGCATCGCGGGGCTTCCATTCAGACTAGGTGTCAATGACTGACAGGGAGGCCTTTGGTTGAGGGCAA
GCCCACGGGGAACTGCAGATGGATGGAAGGGCTCTCCCTGAAGGCTGAGAGGAAGAGTGCAGTCAATTGCAGCCAGT
CCTGCTGGAGCCCAACTTTCTAGAGCCCAGCCCGGCCTTCCCACTCTGTTAACTGCTGGATCGGCTAACCAGGCCGG
TCTCCAGGGCCTTTCAAACACTTACCCAGCCTTTGCCGGCCGTCTTACCATTGCTTGCGTGCGTGTTCATCCC

<210> SEQ ID NO 162

<211> Length : 404

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 162

>M85491_PEA_1_node_8

TGGGAAGAGGTGAGTGGCTACGATGAGAACATGAACACGATCCGCACGTACCAGGTGTGCAACGTGTTTGAGTCAAG
CCAGAACAACTGGCTACGGACCAAGTTTATCCGGCGCCGTGGCGCCCACCGCATCCACGTGGAGATGAAGTTTTCGG
TGCGTGACTGCAGCAGCATCCCCAGCGTGCCTGGCTCCTGCAAGGAGACCTTCAACCTCTATTACTATGAGGCTGAC
TTTGACTCGGCCACCAAGACCTTCCCCAACTGGATGGAGAATCCATGGGTGAAGGTGGATACCATTGCAGCCGACGA
GAGCTTCTCCCAGGTGGACCTGGGTGGCCGCGTCATGAAAATCAACACCGAGGTGCGGAGCTTCGGACCTGTGTCCC
GCAGCGGCTTCTACCTGGC

<210> SEQ ID NO 163

<211> Length : 184

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 163
>M85491_PEA_1_node_9
CTTCCAGGACTATGGCGGCTGCATGTCCCTCATCGCCGTGCGTGTCTTCTACCGCAAGTGCCCCCGCATCATCCAGA
ATGGCGCCATCTTCCAGGAAACCCTGTCGGGGGCTGAGAGCACATCGCTGGTGGCTGCCCGGGGCAGCTGCATCGCC
AATGCGGAAGAGGTGGATGTACCCATCAAG

<210> SEQ ID NO 164
<211> Length : 97
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 164
>M85491_PEA_1_node_10
CTCTACTGTAACGGGGACGGCGAGTGGCTGGTGCCCATCGGGCGCTGCATGTGCAAAGCAGGCTTCGAGGCCGTTGA
GAATGGCACCGTCTGCCGAG

<210> SEQ ID NO 165
<211> Length : 91
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 165
>M85491_PEA_1_node_18
GTTGTCCATCTGGGACTTTCAAGGCCAACCAAGGGGATGAGGCCTGTACCCACTGTCCCATCAACAGCCGGACCACT
TCTGAAGGGGCCAC

<210> SEQ ID NO 166

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 166
>M85491_PEA_1_node_19
CAACTGTGTCTGCCGCAATGGCTACTACAGAGCAGACCTGGACCCCCTGGACATGCCCTGCACAA

<210> SEQ ID NO 167

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 167
>M85491_PEA_1_node_6
AAACGCTAATGGACTCCACTACAGCGACTGCTGAGCTGGGCTGGATGGTGCATCCTCCATCAGGG

<210> SEQ ID NO 168

<211> Length : 810

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 168
>T39971_node_0
GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC

TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGG

<210> SEQ ID NO 169

<211> Length : 168

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 169
>T39971_node_18
GTGCCAGGGGCTGTGGGCCAGGGTAGAAAGCATCTAGGGAGGGTTTGAGAGCTATTGCTCCCAGGGACAGGGTGGAC
AGGGAAGCTGGACCCAGGGCCCTGCAGGACCTGGTGGGAGCTCTGTGAGCACAGGGCAGCCCCAAGACTCCAGGTCC
TGGGCAGTGAACCT

<210> SEQ ID NO 170

<211> Length : 157

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 170
>T39971_node_21
GGTAGTCAGTACTGGCGCTTTGAGGATGGTGTCCTGGACCCTGATTACCCCCGAAATATCTCTGACGGCTTCGATGG
CATCCCGGACAACGTGGATGCAGCCTTGGCCCTCCCTGCCCATAGCTACAGTGGCCGGGAGCGGGTCTACTTCTTCA
AGG

<210> SEQ ID NO 171

<211> Length : 198

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 171

>T39971_node_22

AGTACTCAGGGGGTGGTGGGAGACTGAGCAGGCAGTGGAGCAGTCTTGGATTCCTTTCACATTTCACTGGGGACAGG

CCTCAGCATGTGCCCACCCCTGACCCCCACCTCATGCTGGGAGATCCTAACTTCAACAGCCTCTGGGATCTCCAGTC

TTGCCCTGGCCCAGCCCTCCTAATGCCCACCACCCCGCTCCTCG

<210> SEQ ID NO 172

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 172

>T39971_node_23

GGAAACAGTACTGGGAGTACCAGTTCCAGCACCAGCCCAGTCAGGAGGAGTGTGAAGGCAGCTCCCTGTCGGCTGTG

TTTGAACACTTTGCCATGATGCAGCGGGACAGCTGGGAGGACATCTTCGAGCTTCTCTTCTGGGGCAGAACCTCTG

<210> SEQ ID NO 173

<211> Length : 140

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 173

>T39971_node_31
GCCATCCCGCGCCACGTGGCTGTCCTTGTTCTCCAGTGAGGAGAGCAACTTGGGAGCCAACAACTATGATGACTACA
GGATGGACTGGCTTGTGCCTGCCACCTGTGAACCCATCCAGAGTGTCTTCTTCTTCTCTGGAG

<210> SEQ ID NO 174

<211> Length : 127

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 174
>T39971_node_33
ACAAGTACTACCGAGTCAATCTTCGCACACGGCGAGTGGACACTGTGGACCCTCCCTACCCACGCTCCATCGCTCAG
TACTGGCTGGGCTGCCCAGCTCCTGGCCATCTGTAGGAGTCAGAGCCCAC

<210> SEQ ID NO 175

<211> Length : 223

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 175
>T39971_node_7
TGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAGGAGAAAAACAATGCC
ACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGGGAATCCTGAGCAGAC
ACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGGGGATAGA

<210> SEQ ID NO 176

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 176
>T39971_node_1
CTGACCAAG

<210> SEQ ID NO 177

<211> Length : 44

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 177
>T39971_node_10
GGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG

<210> SEQ ID NO 178

<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 178
>T39971_node_11
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCT

<210> SEQ ID NO 179

<211> Length : 14

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 179
>T39971_node_12
CTTTGCCTTCCGAG

<210> SEQ ID NO 180

<211> Length : 32

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 180
>T39971_node_15
GGCAGTACTGCTATGAACTGGACGAAAAGGCA

<210> SEQ ID NO 181

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 181
>T39971_node_16
GTGAGGCCTGGGTACCCCAAGCTC

<210> SEQ ID NO 182

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 182
>T39971_node_17
ATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCACCCGCATCAACTGTCAGGGGAAGACCTACCT
CTTCAAG

<210> SEQ ID NO 183

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 183
>T39971_node_26
CTGGTACCAGACAGCCCCAGTTCATTAGCCGGGACTGGCACG

<210> SEQ ID NO 184

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 184
>T39971_node_27
GTGTGCCAGGGCAAGTGGACGCAGCCATGGCTGGCCGCATCTACATCTCAG

<210> SEQ ID NO 185

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 185

>T39971_node_28

GCATGGCAC

<210> SEQ ID NO 186

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 186

>T39971_node_29

CCCGCCCCTCCTTGGCCAAGAAACAAAGGTTTAGGCATCGCAACCGCAAAGGCTACCGTTCACAACGAGGCCACAGC
CGTGGCCGCAACCAGAAC

<210> SEQ ID NO 187

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 187

>T39971_node_3

AGTCATGCAAGGGCCGCTGCACTGAGGGCTTCAACGTGGACA

<210> SEQ ID NO 188

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 188
>T39971_node_30
TCCCGCCG

<210> SEQ ID NO 189

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 189
>T39971_node_34
ATGGCCG

<210> SEQ ID NO 190

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 190
>T39971_node_35

GGCCCTCTGTAGCTCCC

<210> SEQ ID NO 191

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 191
>T39971_node_36
TCCTCCCATCTCCTTCCCCCAGCCCAATAAAGGTCCCTTAGCCCCGAAAAAAAAAGCKATAAT

<210> SEQ ID NO 192

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 192
>T39971_node_4
AGAAGTGCCAGTGTGACGAG

<210> SEQ ID NO 193

<211> Length : 58

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 193
>T39971_node_5
CTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTGAGTGCAAGCCCCAAG

<210> SEQ ID NO 194
<211> Length : 6
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 194
>T39971_node_8
CTCAAG

<210> SEQ ID NO 195
<211> Length : 20
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 195
>T39971_node_9
GCCTGAGACCCTTCATCCAG

<210> SEQ ID NO 196
<211> Length : 327

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 196

>Z21368_PEA_1_node_0

AGGTTACTTGACTGGGAGTTCTCAGACCTCCAGTTTCAGCCCTGCCCTCAGCCTCCAATCCGTAAGAGACACCCAGC
CCCAGCAATTGGATTGGGCAGCCCGTCTTGACACACCACTGTGCTGAGTGCTTGAGGACGTGTTTCAACAGATGGTT
GGGGTTAGTGTGTGTCATCACATTCGAGTGGGGATTAAGAGAAGGAAGGCTGCCTTGCTGGAGCTGTGTGGTCTTCT
CCAAGTGAGAGTCGCAGGCAATAGAACTACTTTGCTTTTGGAGGAAAAGGAGGAATTCATTTTCAGCAGACACAAGA
AAAGCAGTTTTTTTTTCAG

<210> SEQ ID NO 197

<211> Length : 177

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 197

>Z21368_PEA_1_node_15

AACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATGAAGTATTCTTGCTG
TGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGTTCAGAG
GACGGATACAGCAGGAACGAAAA

<210> SEQ ID NO 198

<211> Length : 240

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 198
>Z21368_PEA_1_node_19
GGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATCAATGCCTTTGTGACT
ACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAATGTCTACACCAACAA
CGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATCTTAACAACACTGGCT
ACAGAACAG

<210> SEQ ID NO 199
<211> Length : 300
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 199
>Z21368_PEA_1_node_2
TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTT

<210> SEQ ID NO 200
<211> Length : 152
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 200
>Z21368_PEA_1_node_21
CCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGGTGGCGAGAATGGCTTGGATTAATC
AAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAAGCATGGATTTGATTATGCAAAG

<210> SEQ ID NO 201

<211> Length : 176

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 201
>Z21368_PEA_1_node_33
CTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCATTTACACCGCCGACCATGGTTACCATAT
TGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTC
CAAGTGTAGAACCAGGATCAAT

<210> SEQ ID NO 202

<211> Length : 129

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 202
>Z21368_PEA_1_node_36
AGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATG
TGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAG

<210> SEQ ID NO 203

<211> Length : 279

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 203
>Z21368_PEA_1_node_37
GTGTGTCATTGTTCCTCCTCTCAGCCAGCCCCAAATACACTGAGCTCCAGCTGGTGCCCAGAGCCAGCCAGCAGCTG
AAGACATGGAGGCAGAATATGCCTTGCCCACAAGGATCACCCCAAGCTGAGCATTTCTCAGCTGCTTGTGAATAGCA
TATTGATGGAGATGCACTCATGGTCTGTGGGAAGTGAGAGGTGTTTCTTTAAATAAGCTGTTAGCACAGATCCATTT
GGAAAAACGTCCAGATGCCAAAAGTAAATATTATCATTTTGCTTTCAG

<210> SEQ ID NO 280

<211> Length : 853

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 280
>Z21368_PEA_1_node_39
GTAATTATTGGTTCCTGGGGTGCTTCTGGGAACCAGTCCTAGTGGGCAGCTTTCCCTGCTGAGTATTTTTTTTCTCC
TTATTTTTGTTTACTAAGCATGCAGATTTCGTAAACCTAGTCACAAGATTGAATGGTTTGCTGCTTATTCTGTAGTG
GTCAATAGAGTAATAATTGCTGGATCAGAATTGTAAAGAATAACCCTCAAGTTGGTTAATTGGTACAAAAACACAGT
TAGATAGAAGTTATAGAATTTGATAGTATAGTTGGGACATTATCGTTAACAATAATTTATGTATATCTTAAAATAGC
TAGAAGTGAAGAATTGCAAAGTTCCCAACACAAGGAAAAGATAAATGAGATGATGAATATCCCAATTATCTTGATTT
GATCATTACACATTGTAGACTGGTATCCATATATCACACGTACCCCCAAAATATGTATAATTGTGATATATCAATTT
TTAAAATACCAAAAAAGCAAGAGAATGACGACTCCACATCCCCCAAAAAGAATAAATTCTCATAAGCTTGGACCAAA
GCCTTTATCATGGGTGTAGATTTACTGTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTC
TACCATCCTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCTTCTTCCTACTTCCTACTTCTTT
TCCTTGTGTGCCTTTCCTAGTTTTAAATAGATAAATGTATGCCATTGTAATTATTTCCATTGTCACTTCTGGGTTTC
CCCTTTTGGTTCATTAATACCCATTGCCTTGTTTTTCTCTGTACATAAATTAGGAGAGAGAAAATATTTGTATAATT
TTTTTA

<210> SEQ ID NO 281

<211> Length : 162

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 281

>Z21368_PEA_1_node_4

GGATTCTTCACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTAT
TTAGTCTACAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGA
ATGGAGAG

<210> SEQ ID NO 282

<211> Length : 130

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 282

>Z21368_PEA_1_node_41

CAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCA
AAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAG

<210> SEQ ID NO 283

<211> Length : 217

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 283

>Z21368_PEA_1_node_43

AAGTGGCAATGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGT
CCGGCAGAGCACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTT
ACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAA

<210> SEQ ID NO 284

<211> Length : 256

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 284
>Z21368_PEA_1_node_45
AGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATA
AATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGG
GCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCAC
CTACCACTGTCCGAGTGACACACAA

<210> SEQ ID NO 285

<211> Length : 176

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 285
>Z21368_PEA_1_node_53
GGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGAGGAAGG
AGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAACCACTGG
CAGACAGCCCCGTTCTGGAACC

<210> SEQ ID NO 286

<211> Length : 143

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 286

>Z21368_PEA_1_node_56

TGGGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAAT

TTTCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAG

<210> SEQ ID NO 287

<211> Length : 124

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 287

>Z21368_PEA_1_node_58

CTCACAAATACAGTGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTG

TCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATCTTGATGTTG

<210> SEQ ID NO 288

<211> Length : 588

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 288

>Z21368_PEA_1_node_66

AGGACAGTTATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAG

CTACACAGTGTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTT

GAAGGATTTAGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTC

AAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGAC

CCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAAC
ATTAAGTCCAGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAA
AAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTA
GAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAG

<210> SEQ ID NO 289

<211> Length : 585

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 289

>Z21368_PEA_1_node_67

AACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATT
TCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAA
GGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGA
TGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGC
CACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACG
GAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGA
TTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGACATAAGTATATACATGT
TATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAA

<210> SEQ ID NO 290

<211> Length : 1,188

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 290

>Z21368_PEA_1_node_69
TTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAA
ATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGAC
AAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATG
TTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAA
TAAATAAGTAAACAAAATGAAGATTGCCTGCTCTCTCTGTGCCTAGCCTCAAAGCGTTCATCATACATCATACCTTT
AAGATTGCTATATTTTGGGTTATTTTCTTGACAGGAGAAAAAGATCTAAAGATCTTTTATTTTCATCTTTTTTGGTT
TTCTTGGCATGACTAAGAAGCTTAAATGTTGATAAAATATGACTAGTTTTGAATTTACACCAAGAACTTCTCAATAA
AAGAAAATCATGAATGCTCCACAATTTCAACATACCACAAGAGAAGTTAATTTCTTAACATTGTGTTCTATGATTAT
TTGTAAGACCTTCACCAAGTTCTGATATCTTTTAAAGACATAGTTCAAAATTGCTTTTGAAAATCTGTATTCTTGAA
AATATCCTTGTTGTGTATTAGGTTTTTAAATACCAGCTAAAGGATTACCTCACTGAGTCATCAGTACCCTCCTATTC
AGCTCCCCAAGATGATGTGTTTTTGCTTACCCTAAGAGAGGTTTTCTTCTTATTTTTAGATAATTCAAGTGCTTAGA
TAAATTATGTTTTCTTTAAGTGTTTATGGTAAACTCTTTTAAAGAAAATTTAATATGTTATAGCTGAATCTTTTTGG
TAACTTTAAATCTTTATCATAGACTCTGTACATATGTTCAAATTAGCTGCTTGCCTGATGTGTGTATCATCGGTGGG
ATGACAGAACAAACATATTTATGATCATGAATAATGTGCTTTGTAAAAAGATTTCAAGTTATTAGGAAGCATACTCT
GTTTTTTAATCATGTATAATATTCCATGATACTTTTATAGAACAATTCTGGCTTCAGGAAAGTCTAGAAGCAATATT
TCTTCAAATAAAAGGTGTTTAAACTTTTTTCTG

<210> SEQ ID NO 291

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 291
>Z21368_PEA_1_node_11
GACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTG

<210> SEQ ID NO 292

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 292
>Z21368_PEA_1_node_12
ATTATTCAACCAGGATACCTAATTCAAG

<210> SEQ ID NO 293

<211> Length : 15

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 293
>Z21368_PEA_1_node_16
AACATCCGACCCAAC

<210> SEQ ID NO 294

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 294
>Z21368_PEA_1_node_17
ATTATTCTTGTGCTTACCGATGATCAAGATGTGGAGCTGG

<210> SEQ ID NO 295

<211> Length : 74

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 295
>Z21368_PEA_1_node_23
GACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTAAGAGAATGTATCCCCATAG

<210> SEQ ID NO 296

<211> Length : 96

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 296
>Z21368_PEA_1_node_24
GCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTCTAAACTGTACC
CCAATGCTTCCCAACACAT

<210> SEQ ID NO 297

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 297
>Z21368_PEA_1_node_30
AACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTATGCAGTACACAG

<210> SEQ ID NO 298

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 298
>Z21368_PEA_1_node_31
GACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTGATGTCAGTGGAT
GATTCTGTGGAGAGG

<210> SEQ ID NO 299

<211> Length : 57

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 299
>Z21368_PEA_1_node_38
GTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGG

<210> SEQ ID NO 300

<211> Length : 97

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 300
>Z21368_PEA_1_node_47
GTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATA
AGGCATACATTGACAAAGAG

<210> SEQ ID NO 301

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 301
>Z21368_PEA_1_node_49
ATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATG
TAGCTGCAGTAAACAAAG

<210> SEQ ID NO 302

<211> Length : 66

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 302
>Z21368_PEA_1_node_51
CTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCCATTCAA

<210> SEQ ID NO 303

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 303
>Z21368_PEA_1_node_61
GAAATAAAGATGGAGGAAGCTATGACCTACACAG

<210> SEQ ID NO 304

<211> Length : 53

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 304
>Z21368_PEA_1_node_68
CTTGACACCCCTGGTAAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAG

<210> SEQ ID NO 305

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 305
>Z21368_PEA_1_node_7
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAG

SEQ ID NO: 306
>H53626_PEA_1_node_25
GTGCGCAGCGACGTGAAGCCGGTGATCCAGTGGCTGAAGCGCGTGGAGTACGGCGCCGAGGGCCGCCACAACTCCAC
CATCGATGTGGGCGGCCAGAAGTTTGTGGTGCTGCCCACGGGTGACGTGTGGTCGCGGCCCGACGGCTCCTACCTCA
ATAAGCTGCTCATCACCCGTGCCCGCCAGGACGATGCGGGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTAC
AGCTTCCGCAGCGCCTTCCTCACCGTGCTGCCAG

SEQ ID NO: 307
>H53626_PEA_1_node_26
GTGCGCGGCTGCCACGCCACGCCACACCATGCTGGTGCCCGGACCCGCCCCCTGGGCCCGGCGTCCCACCCACCGGG
TGGGGCCCCACCCTTCCCTCCCGGGCCGTGCTGGCCAGGTCATCTGCCGAGGGAGGGCAGCCCAGGGGCACCGTCTC
CACAGCCCCTGGGATGGGTCTGGGGTGCTCTCCTGGTCTTTGTGTCGGCGTTCCCCTCCCTACCTCCTTTCCTCTCG
CTCTTGCAG

SEQ ID NO: 308
>H53626_PEA_1_node_27
ACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTCGGCCACTAGCCTGCCGTGGCCCGTGGTCATCGGCATC
CCAGCCGGCGCTGTCTTCATCCTGGGCACCCTGCTCCTGTGGCTTTGCCAGGCCCAGAAGAAGCCGTGCACCCCCGC
GCCTGCCCCTCCCCTGCCTGGGCACCGCCCGCCGGGGACGGCCCGCGACCGCAGCGGAGACAAGGACCTTCCCTCGT
TGGCCGCCCTCAGCGCTGGCCCTGGTGTGGGGCTGTGTGAGGAGCATGGGTCTCCGGCAGCCCCCCAGCACTTACTG
GGCCCAGGCCCAGTTGCTGGCCCTAAGTTGTACCCCAAACTCTACACAGACATCCACACACACACACACACACACTC
TCACACACACTCACACGTGGAGGGCAAGGTCCACCAGCACATCCACTATCAGTGCTAGACGGCACCGTATCTGCAGT
GGGCACGGGGGGGCCGGCCAGACAGGCAGACTGGGAGGATGGAGGACGGAGCTGCAGACGAAGGCAGGGGACCCATG
GCGAGGAGGAATGGCCAGCACCCCAGGCAGTCTGTGTGTGAGGCATAGCCCCTGGACACACACACACAGACACACAC
ACTACCTGGATGCATGTATGCACACACATGCGCGCACACGTGCTCCCTGAAGGCACACGTACGCACACACGCACATG
CACAGATATGCCGCCTGGGCACACAGATAAGCTGCCCAAATGCACGCACACGCACAGAGACATGCCAGAACATACAA
GGACATGCTGCCTGAACATACACACGCACACCCATGCGCAGATGTGCTGCCTGGACACACACACACACACGGATATG
CTGTCTGGACGCACACACGTGCAGATATGGTATCCGGACACACACGTGCACAG

SEQ ID NO: 309
>H53626_PEA_1_node_34

GCAGATATGCTGCCTGGACACACACACAGATAATGCTGCCTCAACACTCACACACGTGCAGATATTGCCTGGACACA
CACATGTGCACAGATATGCTGTCTGGACATGCACACACGTGCAGATATGCTGTCCGGATACACACGCACGCACACAT
GCAGATATGCTGCCTGGGCACACACTTCCGGACACACATGCACACACAGGTGCAGATATGCTGCCTGGACACACGCA
GACTGACGTGCTTTTGGGAGGGTGTGCCGTGAAGCCTGCAGTACGTGTGCCGTGAGGCTCATAGTTGATGAGGGACT
TTCCCTGCTCCACCGTCACTCCCCCAACTCTGCCCGCCTCTGTCCCCGCCTCAGTCCCCGCCTCCATCCCCGCCTCT
GTCCCCTGGCCTTGGCGGCTATTTTTGCCACCTGCCTTGGGTGCCCAGGAGTCCCCTACTGCTGTGGGCTGGGGTTG
GGGGCACAG

SEQ ID NO: 310
>H53626_PEA_1_node_35
CAGCCCCAAGCCTGAGAGGCTGGAGCCCATGGCTAGTGGCTCATCCCCACTGCATTCTCCCCCTGACACAGAGAAGG
GGCCTTGGTATTTATATTTAAGAAATGAAGATAATATTAATAATGATGGAAGGAAGACTGGGTTGCAGGGACTGTGG
TCTCTCCTGGGGCCCGG

SEQ ID NO: 311
>H53626_PEA_1_node_36
GACCCGCCTGGTCTTTCAGCCATGCTGATGACCACACCCCGTCCAGGCCAGACACCACCCCCCACCCCACTGTCGTG
GTGGCCCCAGATCTCTGTAATTTTATGTAGAGTTTGAGCTGAAGCCCCGTATATTTAATTTATTTTGTTAAACATGA
AAGTGCATCCTTTCCCTCCA

SEQ ID NO: 312
>H53626_PEA_1_node_11
GTCCGGACAGGCCGAGATGACGCCGAGCCCCCTGTTGCTGCTCCTGCTGCCGCCG

SEQ ID NO: 313
>H53626_PEA_1_node_12
CTGCTGCTGGGGGCCTTCCCGCCGGCCGCCGCCGCCCGAG

SEQ ID NO: 314
>H53626_PEA_1_node_16
GTCAACTACACCCTCGTCGTGCTGG

SEQ ID NO: 315
>H53626_PEA_1_node_19
ATGACATTAGCCCAGGGAAGGAGAGCCTGGGGCCCGACAGCTCCTCTGGGG

SEQ ID NO: 316
>H53626_PEA_1_node_20

GTCAAGAGGACCCCGCCAGCCAGCAGTGGG

SEQ ID NO: 317
>H53626_PEA_1_node_24
AGCGGACCCGTTCCAAGCCCGTGCTCACAGGCACGCACCCCGTGAACACGACGGTGGACTTCGGGGGGACCACGTCC
TTCCAGTGCAAG

SEQ ID NO: 318
>H53626_PEA_1_node_28
ATATGCTGCCTGGACACACAGATAATGCTGCCTTGACACACACATGCACGGATATTGCCTGGACACACACACACACA
C

SEQ ID NO: 319
>H53626_PEA_1_node_29
GCGTGCACAGATATGCTGTCTGGACACGCACACACATGCAGATATGCTGCCTGGACACACACTTCCAGACACACGTG
CACAGGCGCAGAT

SEQ ID NO: 320
>H53626_PEA_1_node_30
ATGCTGCCTGGACACACGCAGATATGCTGTCTAGTCACACACACAC

SEQ ID NO: 321
>H53626_PEA_1_node_31
GCAGACATGCTGTCCGGACACACACAC

SEQ ID NO: 322
>H53626_PEA_1_node_32
GCATGCACAGATATGCTGTCCGGACACAC

SEQ ID NO: 323
>H53626_PEA_1_node_33
ACACGCAC

SEQ ID NO: 324
>H53626_PEA_1_P4
MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPPLTMWTKDGRTIHSGWSRF
RVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFTQPSK

MRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAIN
ATYKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGRHNSTIDVGGQKFVVLPT
GDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRSAFLTVLPGARLPRHATPCWCPDPPPGPGVPPTGWG
PTLPSRAVLARSSAEGGQPRGTVSTAPGMGLGCSPGLCVGVPLPTSFPLALADPKPPGPPVASSSSATSLPWPVVIG
IPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPPGTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHL
LGPGPVAGPKLYPKLYTDIHTHTHTHSHTHSHVEGKVHQHIHYQC

SEQ ID NO: 325
>H53626_PEA_1_P5

MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPPLTMWTKDGRTIHSGWSRF
RVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPASQQWARPRFTQPSK
MRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAIN
ATYKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCKTQNRQGHLWPPRPRPLACRGPWSSASQPALSSSWAPCSC
GFARPRRSRAPPRLPLPCLGTARRGRPATAAETRTFPRWPPSALALVWGCVRSMGLRQPPSTYWAQAQLLALSCTPN
STQTSTHTHTHTLTHTHTWRARSTSTSTISARRHRICSGHGGAGQTGRLGGWRTELQTKAGDPWRGGMASTPGSLCV
RHSPWTHTHRHTHYLDACMHTHARTRAP

SEQ ID NO: 326
>Ubiquitin Forward primer
ATTTGGGTCGCGGTTCTTG

SEQ ID NO: 327
>Ubiquitin Reverse primer
TGCCTTGACATTCTCGATGGT

SEQ ID NO: 328
>Ubiquitin-amplicon
ATTTGGGTCGCGGTTCTTGTTTGTGGATCGCTGTGATCGTCACTTGACAATGCAGATCTTCGTGAAGACTCTGACTG
GTAAGACCATCACCCTCGAGG TTGAGCCCAGTGACACCATCGAGAATGTCAAGGCA

SEQ ID NO: 329
>SDHA Forward primer
TGGGAACAAGAGGGCATCTG

SEQ ID NO: 330
SDHA Reverse primer

CCACCACTGCATCAAATTCATG

SEQ ID NO: 331
SDHA-amplicon
TGGGAACAAGAGGGCATCTGCTAAAGTTTCAGATTCCATTTCTGCTCAGTATCCAGTAGTGGATCATGAATTTGATG
CAGTGGTGG

SEQ ID NO: 332
PBGD Forward primer
TGAGAGTGATTCGCGTGGG

SEQ ID NO: 333
PBGD Reverse primer
CCAGGGTACGAGGCTTTCAAT

SEQ ID NO: 334
PBGD-amplicon
TGAGAGTGATTCGCGTGGGTACCCGCAAGAGCCAGCTTGCTCGCATACAGACGGACAGTGTGGTGGCAACATTGAAA
GCCTCGTACCCTGG

<210> SEQ ID NO 335
<211> Length : 760
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 335
>HUMGRP5E_node_0
CCAAAATCTATGGGCTGGGACAGCAAAGATGTGGCCTACGAAGAGAAAGGTCTGGAGAATCAGAAGGCCTTCAAATG
GTGGTTCCAAATCCCTCCAGCAAAGCCCATCCATCTTTAGAGCTCACCCGTCTCCAGCTACACCCCCCACCCCTCCC
GGCCCAGATCAGGCAGCGGGGTCGCCCTCTCCAGGACTCTCAAGGCAGCTAAGGCTGGAGGCGCCGGCGAGCCTGGA
GAGGGAGGAGTTCACTAAATTGTGTTGGATGGAAGGCGTCGAGGACCGGAGGAATTAATCCGATGTGGGGAAGGCGG
ACGGGGCTACGAGGAAAAAAGAGGGGGCAATGTACACTCAGCCTTTTCATCACTCGGCGGGGAGATGGATGGTTTTC
CGGACCGGGCGTCCCAGCGCCCCGGTTAGCTATAGGGAGACGTCAGAGCGCTCTGGTCCGCGATAGAAGAGCCCCCC
AGCCCCCCCGCCCGGGCTTCCATATAAAGTAGGGGCCCTAGTGGAGGCCGCAGCAGTAGCACCAGCGGCTGCGGCGG
CGGAGCTCCTCCGAGGTCCGGGTCACCAGTCTCTGCTCTTCCCAGCCTCTCCGGCGCGCTCCAAGGGCTTCCCGTCG

GGACCATGCGCGGCAGTGAGCTCCCGCTGGTCCTGCTGGCGCTGGTCCTCTGCCTGGCGCCCCGGGGGCGAGCGGTC
CCGCTGCCTGCGGGCGGAGGGACCGTGCTGACCAAGATGTACCCGCGCGGCAACCACTGGGCGGTGG

<210> SEQ ID NO 336

<211> Length : 224

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 336
>HUMGRP5E_node_2
GGCACTTAATGGGGAAAAAGAGCACAGGGGAGTCTTCTTCTGTTTCTGAGAGAGGGAGCCTGAAGCAGCAGCTGAGA
GAGTACATCAGGTGGGAAGAAGCTGCAAGGAATTTGCTGGGTCTCATAGAAGCAAAGGAGAACAGAAACCACCAGCC
ACCTCAACCCAAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAGGATAGCAGCAACTTCAAAGAT

<210> SEQ ID NO 337

<211> Length : 359

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 337
>HUMGRP5E_node_8
GTTCTCAACGTGAAGGAAGGAACCCCCAGCTGAACCAGCAATGATAATGATGGCCTCTCTCAAAAGAGAAAAACAAA
ACCCCTAAGAGACTGCGTTCTGCAAGCATCAGTTCTACGGATCATCAACAAGATTTCCTTGTGCAAAATATTTGACT
ATTCTGTATCTTTCATCCTTGACTAAATTCGTGATTTTCAAGCAGCATCTTCTGGTTTAAACTTGTTTGCTGTGAAC
AATTGTCGAAAAGAGTCTTCCAATTAATGCTTTTTTATATCTAGGCTACCTGTTGGTTAGATTCAAGGCCCCGAGCT
GTTACCATTCACAATAAAAGCTTAAACACATTGTCCAAAGGGCAGGCTGTT

<210> SEQ ID NO 338

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 338
>HUMGRP5E_node_3
GTAGGTTCAAAAGGCAAAG

<210> SEQ ID NO 339

<211> Length : 14

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 339
>HUMGRP5E_node_7
ACTCTCTGCTCCAG

<210> SEQ ID NO 340

<211> Length : 178

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 340
>D56406_PEA_1_node_0
TTCACTCACTTTCAAAGCCAGCTGAAGGAAAGAGGAAGTGCTAGAGAGAGCCCCCTTCAGTGTGCTTCTGACTTTTA
CGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGGAATGAAAATCCAGCTTGTATGCATGCTACTCCTGGCTT
TCAGCTCCTGGAGTCTGTGCTCAG

<210> SEQ ID NO 341

<211> Length : 780

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 341

>D56406_PEA_1_node_13

TTAATCCAGGAAGATATTCTTGATACTGGAAATGACAAAAATGGAAAGGAAGAAGTCATAAAGAGAAAAATTCCTTA
TATTCTGAAACGGCAGCTGTATGAGAATAAACCCAGAAGACCCTACATACTCAAAAGAGATTCTTACTATTACTGAG
AGAATAAATCATTTATTTACATGTGATTGTGATTCATCATCCCTTAATTAAATATCAAATTATATTTGTGTGAAAAT
GTGACAAACACACTTATCTGTCTCTTCTACAATTGTGGTTTATTGAATGTGATTTTTCTGCACTAATATAAATTAGA
CTAAGTGTTTTCAAATAAATCTAAATCTTCAGCATGATGTGTTGTGTATAATTGGAGTAGATATTAATTAAGTCACC
TGTATAATGTTTTGTAATTTTGCAAAACATATCTTGAGTTGTTTAAACAGTCAAAATGTTTGATATTTTATACCAGC
TTATGAGCTCAAAGTACTACAGCAAAGCCTAGCCTGCATATCATTCACCCAAAACAAAGTAATAGCGCCTCTTTTAT
TATTTTGACTGAATGTTTTATGGAATTGAAAGAAACATACGTTCTTTTCAAGACTTCCTCATGAATCTCTCAATTAT
AGGAAAAGTTATTGTGATAAAATAGGAACAGCTGAAAGATTGATTAATGAACTATTGTTAATTCTTCCTATTTTAAT
GAATGACATTGAACTGAATTTTTTGTCTGTTAAATGAACTTGATAGCTAATAAAAAGACAACTAGCCATCAAAATCA
AAAGTTTCTC

<210> SEQ ID NO 342

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 342

>D56406_PEA_1_node_11

GCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGACGGGCGGATCACGAGGTCAAGAGATGGAGAC
CATCCCGGCTAACACG

<210> SEQ ID NO 343

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 344
>D56406_PEA_1_node_2
ATTCAG

<210> SEQ ID NO 344

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 344
>D56406_PEA_1_node_3
AAGAGGAAATGAAAGCATTAGAAGCAGATTTCTTGACCAATATGCATACATCAAAG

<210> SEQ ID NO 345

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 345
>D56406_PEA_1_node_5

ATTAGTAAAGCACATGTTCCCTCTTGGAAGATGACTCTGCTAAATGTTTGCAGTCTTGTAAATAATTTGAACAGCCC
AGCTGAGGAAACAGGAGAAGTTCATGAAGAGGAGCTTG

<210> SEQ ID NO 346

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 346
>D56406_PEA_1_node_6
TTGCAAGAAGGAAACTTCCTACTGCTTTAGATGG

<210> SEQ ID NO 347

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 347
>D56406_PEA_1_node_7
CTTTAGCTTGGAAGCAATGTTGACAA

<210> SEQ ID NO 348

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 348
>D56406_PEA_1_node_8
TATACCAG

<210> SEQ ID NO 349
<211> Length : 42
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 349
>D56406_PEA_1_node_9
CTCCACAAAATCTGTCACAGCAGGGCTTTTCAACACTGGGAG

<210> SEQ ID NO 350
<211> Length : 245
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 350
>F05068_PEA_1_node_0
AAGAAAGGGAAGGCAACCGGGCAGCCCAGGCCCCGCCCCGCCGCTCCCCCACCCGTGCGCTTATAAAGCACAGGAAC
CAGAGCTGGCCACTCAGTGGTTTCTTGGTGACACTGGATAGAACAGCTCAAGCCTTGCCACTTCGGGCTTCTCACTG
CAGCTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACTTTCTCCTTCAAGTACTTGGCAGAT
CACTCTCTTAGCAG

<210> SEQ ID NO 351
<211> Length : 161

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 351

>F05068_PEA_1_node_10

CTTCGGGACGTGCACGGTGCAGAAGCTGGCACACCAGATCTACCAGTTCACAGATAAGGACAAGGACAACGTCGCCC

CCAGGAGCAAGATCAGCCCCCAGGGCTACGGCCGCCGGCGCCGGCGCTCCCTGCCCGAGGCCGGCCCGGGTCGGACT

CTGGTGT

<210> SEQ ID NO 352

<211> Length : 121

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 352

>F05068_PEA_1_node_12

CCATGGTACAAGGAATAGTCGCGCAAGCATCCCGCTGGTGCCTCCCGGGACGAAGGACTTCCCGAGCGGTGTGGGGA

CCGGGCTCTGACAGCCCTGCGGAGACCCTGAGTCCGGGAGGCAC

<210> SEQ ID NO 353

<211> Length : 630

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 353

>F05068_PEA_1_node_13

CGTCCGGCGGCGAGCTCTGGCTTTGCAAGGGCCCCTCCTTCTGGGGGCTTCGCTTCCTTAGCCTTGCTCAGGTGCAA

GTGCCCCAGGGGGCGGGGTGCAGAAGAATCCGAGTGTTTGCCAGGCTTAAGGAGAGGAGAAACTGAGAAATGAATGC

TGAGACCCCCGGAGCAGGGGTCTGAGCCACAGCCGTGCTCGCCCACAAACTGATTTCTCACGGCGTGTCACCCCACC

AGGGCGCAAGCCTCACTATTACTTGAACTTTCCAAAACCTAAAGAGGAAAAGTGCAATGCGTGTTGTACATACAGAG
GTAACTATCAATATTTAAGTTTGTTGCTGTCAAGATTTTTTTTGTAACTTCAAATATAGAGATATTTTTGTACGTTA
TATATTGTATTAAGGGCATTTTAAAAGCAATTATATTGTCCTCCCCCTATTTTAAGACGTGAATGTCTCAGCGAGGT
GTAAAGTTGTTCGCCGCGTGGAATGTGAGTGTGTTTGTGTGCATGAAAGAGAAAGACTGATTACCTCCTGTGTGGAA
GAAGGAAACACCGAGTCTCTGTATAATCTATTTACATAAAATGGGTGATATGCGAACAGCAAACCAATAAACTGTCT
CAATGCTGAATAAAA

<210> SEQ ID NO 354

<211> Length : 150

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 354
>F05068_PEA_1_node_4
GTGAGTCCGGGCAGCGCCTTCCCCCTTGCTGGTACCTGGCAGGCAAGGGGAACTGACCGTTGGTCCCGAAGGTCTAG
AAGTGAATGGGAGCAGGGACAGGCCTGGGCGTCACCTGAACGCACGCGAATCGGGTCTGCTTGTGTTTTCCAG

<210> SEQ ID NO 355

<211> Length : 233

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 355
>F05068_PEA_1_node_8
GTAACTACGCCCTGTGCTGTCCAGGGACGGGAGGGAAGGAAGGTGTGCGGGAGGAGTTCTCTGTCTCCACTCCCCTG
GCCCGGGGGATCGTCGGGGCTGGACCGCAGCTCAGATGGCGCGAGCAGTTTCCAGCTCCCTCTGGCTCTAGAATGGC
TCCCGTTCCCGGTGTTGGGGCCAAAGCTCTGCTTGATGGGGTCTCAAGTTGCCTTTCTTCCCCCTCCCCCCGCCCGC
AG

<210> SEQ ID NO 356

<211> Length : 76

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 356

>F05068_PEA_1_node_11

CTTCTAAGCCACAAGCACACGGGGCTCCAGCCCCCCCGAGTGGAAGTGCTCCCCACTTTCTTTAGGATTTAGGCGC

<210> SEQ ID NO 357

<211> Length : 119

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 357

>F05068_PEA_1_node_3

GGTCTGCGCTTCGCAGCCGGGATGAAGCTGGTTTCCGTCGCCCTGATGTACCTGGGTTCGCTCGCCTTCCTAGGCGC

TGACACCGCTCGGTTGGATGTCGCGTCGGAGTTTCGAAAGAA

<210> SEQ ID NO 358

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 358

>F05068_PEA_1_node_5

GTGGAATAAGTGGGCTCTGAGTCGTGGGAAGAGGGAACTGCGGATGTCCAGCAGCTACC

<210> SEQ ID NO 359

<211> Length : 60

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 359
>F05068_PEA_1_node_6
CCACCGGGCTCGCTGACGTGAAGGCCGGGCCTGCCCAGAC

<210> SEQ ID NO 360

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 360
>F05068_PEA_1_node_7
CCTTATTCGGCCCCAGGACATGAAGGGTGCCTCTCGAAGCCCCGAAGACAG

<210> SEQ ID NO 361

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 361
>F05068_PEA_1_node_9
CAGTCCGGATGCCGCCCGCATCCGAGTCAAGCGCTACCGCCAGAGCATGAACAACTTCCAGGGCCTCCGGAGCTTTG
GCTGCCG

<210> SEQ ID NO 362

<211> Length : 573

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 362

>H14624_node_0

TTATGCTCCCGCGGAGGCCAAGCGGACTCCCTGACAGGACAGAATCTGAACGTGAGAGTGAAGGTCTTGCCTGTCCA
GAAACTCTTGTAGCCAGCACAGGTTTAAACAAGAAGCCAAATTGTTCTGGAGAGATTGCTGGGGGCTTTCTTTGTGC
CTCAAGCTTCTTCAGTGCCCTGAGCACAGGAAACACTCAAGCAGAGAAGCAGAGCCAAACCCAGGATACGGGAGGTC
GAGGCTCTTCCGTAGACCTGCAGCATTGGGGTGGGATGATGTTCATTCTGTGTGTGTTCTGGACCAAGCCCCTCTCC
AGGGACCTATGGGCAGCCCCCTTTAAGCAAGATGCCCGGTGGAGTGGGCATCCACCATCACTTACCCTGGGCTTGGG
TGAATAGATTTTCCGTGCCTTAAATGGGCAGGGAGGGGGTAAACATGGACGGTCCATTGGTACAAATAAAAGCCTTT
GGTGGGTTTTGATCAATTGCAAGGATCGAAGGAGACCTGTGGACCTGAGGTCAACTGGCAGCAGAGAAGAGTCTGGG
TTCGTGAAGGCGCCGCCGCGGTGCCGCGCCACGT

<210> SEQ ID NO 363

<211> Length : 387

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 363

>H14624_node_16

GTAAGCCTTCCCTCTTGCTTCCCCACTCCCTGCTGGGCTGAGACGCTCCCAGGAGATCCCGCCCCTGCCACGCATCC
CAGTGCATCCCTGCTTGGGGTGCCAGTAGCGGGAAGGGCAGAAGTTCTGCCTGACCTGGTCTGTCATCACAACAAGC
CTGTATCAAATTTGAGGCACCCCTCCCACGCCGCCCAAGTCTCGCGCATTCTCCTCCCGAGTTGTACCAGCTATACT
TAAGGGCAGTTTAAAAATAAAACAAACAAACAAAAACAACAAAACTAAAAAAAACGAAGAACTGAACGGCGGTTTAAA
AAAAAATAGATACACGATTATTGTTAAAGATGCTAGCACTGGAGCTGCGCAGAGCGTTGGAAGTGGTGTTTGGTGGA
GG

<210> SEQ ID NO 364

<211> Length : 249

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 364

>H14624_node_3

ATTTGCATAAAAAAGGCCAAGAAAACTCTGGCTGTGCCCCAGCAACGGCTCATTCTGCTCCCCCGGGTCGGAGCCCC

CCGGAGCTGCGCGCGGGCTTGCAGCGCCTCGCCCGCGCTGTCCTCCCGGTGTCCCGCTTCTCCGCGCCCCAGCCGCC

GGCTGCCAGCTTTTCGGGGCCCCGAGTCGCACCCAGCGAAGAGAGCGGGCCCGGGACAAGCTCGAACTCCGGCCGCC

TCGCCCTTCCCCGGCTCC

<210> SEQ ID NO 365

<211> Length : 10

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 365

>H14624_node_10

GTGCTGGAGC

<210> SEQ ID NO 366

<211> Length : 35

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 366

>H14624_node_11

AGGCCGGCGCTTGGATCCCGCTGGTCATGAAGCAG

<210> SEQ ID NO 367

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 367
>H14624_node_12
TGCCACCCGGACACCAAGAAG

<210> SEQ ID NO 368

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 368
>H14624_node_13
TTCCTGTGCTCGCTCTTCGCCCCCGTCTGCCTCGATGACCTAGACGAGACCATCCAGCCATGCCACTCGCTCTGCGT
GCAGGTGAAGGACCG

<210> SEQ ID NO 370

<211> Length : 60

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 370
>H14624_node_14
CTGCGCCCCGGTCATGTCCGCCTTCGGCTTCCCCTGGCCCGACATGCTTGAGTGCGACCG

<210> SEQ ID NO 371

<211> Length : 71

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 371
>H14624_node_15
TTTCCCCCAGGACAACGACCTTTGCATCCCCCTCGCTAGCAGCGACCACCTCCTGCCAGCCACCGAGGAAG

<210> SEQ ID NO 372

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 372
>H14624_node_4
GCTCCCTCTGCCCCCTCGGGGTCGCGCGCCCACGATGCTGCAGGGCCCTGGCTCGCTGCTGCTGCTCTTC

<210> SEQ ID NO 373

<211> Length : 11

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 373
>H14624_node_5
CTCGCCTCGCA

<210> SEQ ID NO 374

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 374
>H14624_node_6
CTGCTGCCTGGGCTCGGCGCGCGG

<210> SEQ ID NO 375

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 375
>H14624_node_7
GCTCTTC

<210> SEQ ID NO 376

<211> Length : 80

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 376
>H14624_node_8
CTCTTTGGCCAGCCCGACTTCTCCTACAAGCGCAGCAATTGCAAGCCCATCCCGGCCAACCTGCAGCTGTGCCACGG
CAT

<210> SEQ ID NO 377

<211> Length : 55

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 377
>H14624_node_9
CGAATACCAGAACATGCGGCTGCCCAACCTGCTGGGCCACGAGACCATGAAGGAG

<210> SEQ ID NO 378

<211> Length : 213

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 378
>H38804_PEA_1_node_0
GACTGGGTTGACCGATGCTGGGCAGCTGAGCGGACCAATCGGCCCCCTAGACTGAGACGTTGGCGTTTGAAATCAGC
CAATGGCAGGTCTACACTGGAGCTTCCTCTCCGCCTCCTTCGCCTAGCCTGCGAGTGTTCTGAGGGAAGCAAGGAGG
CGGCGGCGGCCGCAGCGAGTGGCGAGTAGTGGAAACGTTGCTTCTGAGGGGAGCCCAAG

<210> SEQ ID NO 379

<211> Length : 432

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 379
>H38804_PEA_1_node_1
GTAGGGAGGCGAGGCGACGGTGTGCGGGAGCGGGCTCTCCAGGGACTTCCCGGGTCCGCAACTGGCAGGGCCGTTCG
ATTCGCAGGGGATCCCGTTTCGTTTCTGTTGTTTTCCCTTTATTTTTAGGAGTGCCCGGGGCGACGGGACCCCGGGA

GAGGGGAAAGGGAACAGTCTGGGGTCCGGGCATCGCTGTGGGCCGGGCTGGGTTTAGGGGGACGGCGGTGCGGGCTG
GGCCGGTTTGGGCGCGGCGGGGGCCGGATGATGGGGCGAGTCCGGACCTTGGCGGGCGAGTGCTCGGCGCAGGCGCA
AGCGCAGAGTCTCCTCGCGGTCGTCCTCTCGGCCCCTCCCTCTGGGGGGACCCCCAGTGCCAGGCTGTCAGTGCGCA
GCCCCAGCCCGCGGGACCCCTGGGGACTCTGGGCGCCTGTTCTGCAG

<210> SEQ ID NO 380

<211> Length : 159

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 380
>H38804_PEA_1_node_16
GTATATACCCTCTCAGTGTCTGGAGACCGGCTGATTGTGGGAACAGCAGGCCGCAGAGTGTTGGTGTGGGACTTACG
GAACATGGGTTACGTGCAGCAGCGCAGGGAGTCCAGCCTGAAATACCAGACTCGCTGCATACGAGCGTTTCCAAACA
AGCAG

<210> SEQ ID NO 381

<211> Length : 139

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 381
>H38804_PEA_1_node_19
GGTTATGTATTAAGCTCTATTGAAGGCCGAGTGGCAGTTGAGTATTTGGACCCAAGCCCTGAGGTACAGAAGAAGAA
GTATGCCTTCAAATGTCACAGACTAAAAGAAAATAATATTGAGCAGATTTACCCAGTCAATG

<210> SEQ ID NO 382

<211> Length : 196

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 382
>H38804_PEA_1_node_24
ATATTTGGGATCCATTTAACAAAAAGCGACTGTGCCAATTCCATCGGTACCCCACGAGCATCGCATCACTTGCCTTC
AGTAATGATGGGACTACGCTTGCAATAGCGTCATCATATATGTATGAAATGGATGACACAGAACATCCTGAAGATGG
TATCTTCATTCGCCAAGTGACAGATGCAGAAACAAAACCCAA

<210> SEQ ID NO 383
<211> Length : 353
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 383
>H38804_PEA_1_node_25
GTGAGTATGCTTCACCTGTATTTGAGCCTTTTCTTGCATTCAACCCAGGATTTATTAATTTTTCTAAATTCATGAAT
AGCATTGTTGATGCCTGCTCGATATTACAGCTGACTGTAGGGTTGGAGTTGATGTTATCATGTTCTCCCAAGCTTTC
AATATCCGTAGGTTGATAGACGTCTGATGGATAAAATTGTGCCTAGTTGTTTTGTAGAGAAGAATGTCAAACTCTTA
TTCTTCTTGAATAGGCTCTATTATTTGAATCTCTGGAGTTATTACCAGCTCATTGCTTCAAAATTAAGTTGAGGAAT
TCAAGAATAATTTATTTTAGTAAATTCTATTTAAGATGTTTAAGA

<210> SEQ ID NO 384
<211> Length : 590
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 384
>H38804_PEA_1_node_28
TATTAACACAAAGTAAGTGACCTTCAGGTCTTATTGGAAACTCAGAGTAATATGGCCTTGCCTGGAATTGCAAATTT
CCTTAGTTTTGAAATTTTCATAGATGTCTTTGGTTCTTGGTTGTAACTGTTGACTGAGAAGAGCCATTTACATTTTT
TGATACCAACAGGGCAAAGCTTTTTACTTAATTACCTCTACCAGGCTTTAAGGGAAATCTGATACTTCAGCATGTGT
TAAACTATAAAATACCTACTCCAAGTATCTGCCCAGTTCCTTGTCCCCTCTCCCCAGGCCCTTAAAGGAAGTTCTCG
ATACATATTTGTAGAATAACTGAATGTTTTCAGGATTCCTGTACTTTGCTGAGTTAAAATGGATATGGTACCCTTGC
TGATTGGTTGAGCCCCTAAGAGGGGGCAGAATATTAAATATTCCATATCAGATATGCTTTTACAGGTTTGACTTTAG

AAAAGTCTTAGCATGTGAAGCCTGTTGGATAAAGGGCTGTGTTTGCATTTAATCTGTCACTTTTGTATCTCCTGTCC
TGGCTGGCCATTTTGATCTCATGCTGTTCTTTTTTTCTTTTGAACTTGTAG

<210> SEQ ID NO 385

<211> Length : 1,228

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 385
>H38804_PEA_1_node_29
GTCACCATGTACTTGACAAGATTTCATTTACTTAAGTGCCATGTTGATGATAATAAAACAATTCGTACTCCCCAATG
GTGGATTTATTACTATTAAAGAAACCAGGGAAAATATTAATTTTAATATTATAACAACCTGAAAATAATGGAAAAGA
GGTTTTTGAATTTTTTTTTTTAAATAAACACCTTCTTAAGTGCATGAGATGGTTTGATGGTTTGCTGCATTAAAGGT
ATTTGGGCAAACAAAATTGGAGGGCAAGTGACTGCAGTTTTGAGAATCAGTTTTGACCTTGATGATTTTTTGTTTCC
ACTGTGGAAATAAATGTTTGTAAATAAGTGTAATAAAAATCCCTTTGCATTCTTTCTGGACCTTAAATGGTAGAGGA
AAAGGCTCGTGAGCCATTTGTTTCTTTTGCTGGTTATAGTTGCTAATTCTAAAGCTGCTTCAGACTGCTTCATGAGG
AGGTTAATCTACAATTAAACAATATTTCCTCTTGGCCGTCCATTATTTTCTGAAGCAGATGGTTCATCATTTCCTGG
GCTGTTAAACAAAGCGAGGTTAAGGTTAGACTCTTGGGAATCAGCTAGTTTTCAATCTTATTAGGGTGCAGAAGGAA
AACTAATAAGAAAACCTCCTAATATCATTTTGTGACTGTAAACAATTATTTATTAGCAAACAATTGATCCCAGAAGG
GCAAATTGTTTGAGTCAGTAATGAGCTGAGAAAAGACAGAGCATATCTGTGTATTTGGAAAAATAATTGTAACGTAA
TTGCAGTGCATTTAGACAGGCATCTATTTGGACCTGTTTCTATCTCTAAATGAATTTTTGGAAACATTAATGAGGTT
TACATATTTCTCTGACATTTATATAGTTCTTATGTCCATTTCAGTTGACCAGCCGCTGGTGATTAAAGTTAAAAAGA
AAAAAATTATAGTGAGAATGAGATTCATTTCAATGTAATGCACTAAAGCAGAACACGAACTTAGCTTGGCCTATTCT
AGGTAGTTCCAAATAGTATTTTTGTTGTCAAACTTTAAAATTTATATTAATTTGCAAATGTATGTCTCTGAGTAGGA
CTTGGACCTTTCCTGAGATTTATTTTATCCGTGATGTATTTTTTTTAATTCTTTTGATACAGAGAAGGGTCTTTTTT
TTTTTTAAGTATTTCAGTGAAAACTTGGTGTAAGTCTGAACCCATCTTTTGAAATGTATTTTCTTCATTGCAG

<210> SEQ ID NO 386

<211> Length : 326

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 386

>H38804_PEA_1_node_30
GTCCACCTAATCATCCTGTGAAAGTGGTTTCTCTATGGAAAGCTTTGTTTGCTTCCTACAAATACATGCTTATTCCT
TAAGGGATGTGTTAGAGTTACTGTGGATTTCTCTGTTTTCTGTCTTACAAGAAACTTGTCTATGTACCTTAATACTT
TGTTTAGGATGAGGAGTCTTTGTGTCCCTGTACAGTAGTCTGACGTATTTCCCCTTCTGTCCCCTAGTAAGCCCAGT
TGCTGTATCTGAACAGTTTGAGCTCTTTTTGTAATATACTCTAAACCTGTTATTTCTGTGCTAATAAACGAGATGCA
GAACCCTTGAAAAATGGA

<210> SEQ ID NO 387

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 387
>H38804_PEA_1_node_10
GATCCAACGCATGCCTGGAGTGGAGGACTAGATCATCAATTGAAAATGCATGATTTGAACACTGATCAAG

<210> SEQ ID NO 388

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 388
>H38804_PEA_1_node_12
AAAATCTTGTTGGGACCCATGATGCCCCTATCAGATGTG

<210> SEQ ID NO 389

<211> Length : 79

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 389
>H38804_PEA_1_node_13
TTGAATACTGTCCAGAAGTGAATGTGATGGTCACTGGAAGTTGGGATCAGACAGTTAAACTGTGGGATCCCAGAACT
CC

<210> SEQ ID NO 390

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 390
>H38804_PEA_1_node_14
TTGTAATGCTGGGACCTTCTCTCAGCCTGAAAAG

<210> SEQ ID NO 391

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 391
>H38804_PEA_1_node_2
ATGACCGGTTCTAACGAGTTCAAGCTGAACCAG

<210> SEQ ID NO 392

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 392
>H38804_PEA_1_node_20
CCATTTCTTTTCACAATATCCACAATACATTTGCCACAG

<210> SEQ ID NO 393

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 393
>H38804_PEA_1_node_23
GTGGTTCTGATGGCTTTGTAA

<210> SEQ ID NO 394

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 394
>H38804_PEA_1_node_26
ATTTGAACTGCCAAAAATCTTTCCTCTCCACAGAGGTTGTTTCTTTAA

<210> SEQ ID NO 395

<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 395

>H38804_PEA_1_node_3

CCACCCGAGGATGGCATCTCCTCCGTGAAGTTCAGCCC

<210> SEQ ID NO 396

<211> Length : 111

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 396

>H38804_PEA_1_node_4

CAACACCTCCCAGTTCCTGCTTGTCTCCTCCTGGGACACGTCCGTGCGTCTCTACGATGTGCCGGCCAACTCCATGC

GGCTCAAGTACCAGCACACCGGCGCCGTCCTGGA

<210> SEQ ID NO 397

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 397

>H38804_PEA_1_node_5

CTGCGCCTTCTAC

<210> SEQ ID NO 398

<211> Length : 257

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 398

>HSENA78_node_0

AGTGGGGAGAGATGAGTGTAGATAAAAGGAGTGCAGAAGGCACGAGGAAGCCACAGTGCTCCGGATCCTCCAATCTT

CGCTCCTCCAATCTCCGCTCCTCCACCCAGTTCAGGAACCCGCGACCGCTCGCAGCGCTCTCTTGACCACTATGAGC

CTCCTGTCCAGCCGCGCGGCCCGTGTCCCCGGTCCTTCGAGCTCCTTGTGCGCGCTGTTGGTGCTGCTGCTGCTGCT

GACGCAGCCAGGGCCCATCGCCAGCG

<210> SEQ ID NO 399

<211> Length : 133

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 399

>HSENA78_node_2

CTGGTCCTGCCGCTGCTGTGTTGAGAGAGCTGCGTTGCGTTTGTTTACAGACCACGCAGGGAGTTCATCCCAAAATG

ATCAGTAATCTGCAAGTGTTCGCCATAGGCCCACAGTGCTCCAAGGTGGAAGTGGT

<210> SEQ ID NO 400

<211> Length : 1,786

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 400

>HSENA78_node_6

TGGAAACAAGGAAAACTGATTAAGAGAAATGAGCACGCATGGAAAAGTTTCCCAGTCTTCAGCAGAGAAGTTTTCTG
GAGGTCTCTGAACCCAGGGAAGACAAGAAGGAAAGATTTTGTTGTTGTTTGTTTATTTGTTTTTCCAGTAGTTAGCT
TTCTTCCTGGATTCCTCACTTTGAAGAGTGTGAGGAAAACCTATGTTTGCCGCTTAAGCTTTCAGCTCAGCTAATGA
AGTGTTTAGCATAGTACCTCTGCTATTTGCTGTTATTTTATCTGCTATGCTATTGAAGTTTTGGCAATTGACTATAG
TGTGAGCCAGGAATCACTGGCTGTTAATCTTTCAAAGTGTCTTGAATTGTAGGTGACTATTATATTTCCAAGAAATA
TTCCTTAAGATATTAACTGAGAAGGCTGTGGATTTAATGTGGAAATGATGTTTCATAAGAATTCTGTTGATGGAAAT
ACACTGTTATCTTCACTTTTATAAGAAATAGGAAATATTTTAATGTTTCTTGGGGAATATGTTAGAGAATTTCCTTA
CTCTTGATTGTGGGATACTATTTAATTATTTCACTTTAGAAAGCTGAGTGTTTCACACCTTATCTATGTAGAATATA
TTTCCTTATTCAGAATTTCTAAAAGTTTAAGTTCTATGAGGGCTAATATCTTATCTTCCTATAATTTTAGACATTCT
TTATCTTTTTAGTATGGCAAACTGCCATCATTTACTTTTAAACTTTGATTTTATATGCTATTTATTAAGTATTTTAT
TAGGAGTACCATAATTCTGGTAGCTAAATATATATTTTAGATAGATGAAGAAGCTAGAAAACAGGCAAATTCCTGAC
TGCTAGTTTATATAGAAATGTATTCTTTTAGTTTTTAAAGTAAAGGCAAACTTAACAATGACTTGTACTCTGAAAGT
TTTGGAAACGTATTCAAACAATTTGAATATAAATTTATCATTTAGTTATAAAAATATATAGCGACATCCTCGAGGCC
CTAGCATTTCTCCTTGGATAGGGGACCAGAGAGAGCTTGGAATGTTAAAAACAAAACAAAACAAAAAAAAACAAGGA
GAAGTTGTCCAAGGGATGTCAATTTTTTATCCCTCTGTATGGGTTAGATTTTCCAAAATCATAATTTGAAGAAGGCC
AGCATTTATGGTAGAATATATAATTATATATAAGGTGGCCACGCTGGGGCAAGTTCCCTCCCCACTCACAGCTTTGG
CCCCTTTCACAGAGTAGAACCTGGGTTAGAGGATTGCAGAAGACGAGCGGCAGCGGGGAGGGCAGGGAAGATGCCTG
TCGGGTTTTTAGCACAGTTCATTTCACTGGGATTTTGAAGCATTTCTGTCTGAATGTAAAGCCTGTTCTAGTCCTGG
TGGGACACACTGGGGTTGGGGGTGGGGGAAGATGCGGTAATGAAACCGGTTAGTCAGTGTTGTCTTAATATCCTTGA
TAATGCTGTAAAGTTTATTTTTACAAATATTTCTGTTTAAGCTATTTCACCTTTGTTTGGAAATCCTTCCCTTTTAA
AGAGAAAATGTGACACTTGTGAAAAGGCTTGTAGGAAAGCTCCTCCCTTTTTTTCTTTAAACCTTTAAATGACAAAC
CTAGGTAATTAATGGTTGTGAATTTCTATTTTTGCTTTGTTTTTAATGAACATTTGTCTTTCAGAATAGGATTCTGT
GATAATATTTAAATGGCAAAAACAAAACATAATTTTGTGCAATTAACAAAGCTACTGCAAGAAAAATAAAACATTTC
TTGGTAAAAACGTAT

<210> SEQ ID NO 401

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 401

>HSENA78_node_9

ATATAATATATATTATATATTTAGCATTGCTGAGCTTTTTAGATGCCTATTGTGTATCTTTTAAAGGTTTTGACCAT
TTTGTTATGAGTAATTACATATATATTACATTCACTATATTAAAATTGTACTTTTTTACTATGTGTCTCATTGGTT

<210> SEQ ID NO 402

<211> Length : 110

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 402
>HSENA78_node_3
GTAAGTTCTGTGCTGCTGTGTCCGCTGTGACCTTGGCAAGAGAGAAATCCCGCAGCCTGGGTCTTCAACCTTGGTAT
CTCATGAGTGTATCTTCTTTTTCTTTCCTTCAG

<210> SEQ ID NO 403

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 403
>HSENA78_node_4
AGCCTCCCTGAAGAACGGGAAGGAAATTTGTCTTGATCCAGAAGCCCCTTTTCTAAAGAAAGTCATCCAGAAAATTT
TGGACGG

<210> SEQ ID NO 404

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 404

>HSENA78_node_8
GTATTTATATATTATATATTTAT

<210> SEQ ID NO 405

<211> Length : 139

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 405
>HUMODCA_node_1
GTGCGTCTCCATGGCGACCCGCCGGTGCTATAAGTAGGGAGCGGCGTGCCGTGGGGCTTTGTCAGTCCCTCCTGTAG
CCGCCGCCGCCGCCGCCCGCCGCCCCTCTGCCAGCAGCTCCGGCGCCACCTCGGGCCGGCGT

<210> SEQ ID NO 406

<211> Length : 135

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 406
>HUMODCA_node_25
GTTGGTTTTGCGGATTGCCACTGATGATTCCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCACGCTCAGAA
CCAGCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATATCGATGTTGTTGGTGTCAG

<210> SEQ ID NO 407

<211> Length : 163

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 407
>HUMODCA_node_32
ATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGACTCTGGAGTGAGAATCATAGCTGAGCCCGG
CAGATACTATGTTGCATCAGCTTTCACGCTTGCAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGG
GCTCTGATG

<210> SEQ ID NO 408

<211> Length : 215

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 408
>HUMODCA_node_36
AGACCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTGTTGA
GCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAAACATGGGCGCTTACACTGTTGCTGCTG
CCTCTACGTTCAATGGCTTCCAGAGGCCGACGATCTACTATGTGATGTCAGGGCCTGCGTG

<210> SEQ ID NO 409

<211> Length : 173

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 409
>HUMODCA_node_39
GATGCCAGCACCCTGCCTGTGTCTTGTGCCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAG
TATTAATGTGTAGATAGCACTCTGGTAGCTGTTAACTGCAAGTTTAGCTTGAATTAAGGGATTTGGGGGGACCATGT
AACTTAATTACTGCTAGTT

<210> SEQ ID NO 410

<211> Length : 1,096

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 410

>HUMODCA_node_41

TTGAATATTTGTTTTATATGGATTTTTATTCACTCTTCAGACACGCTACTCAAGAGTGCCCCTCAGCTGCTGAACAA
GCATTTGTAGCTTGTACAATGGCAGAATGGGCCAAAAGCTTAGTGTTGTGACCTGTTTTTAAAATAAAGTATCTTGA
AATAATTAGGCATTGGGACGTTTTTATGGTGTGTTCATTCCAGACAGTTCACGAATCCCGTATAGCTCGCTCTGATT
CTCAGAGAACAATGAGTGGGTCCACCCACACACAGGTAGGAGGACAGGTGAGACGGAAGCCCCATCCTCCCATGTGG
ACGGTGCACATCTGCTCAGCCCACCCCACATGTCCAGAGTTGGCTGCAAACTCCTTGTCCAGAGCCTCTGGTGGTGG
GACCTACTTAAGTCTGACGGACCTGTCCTGTCCAGGCCAGTGCCCAGGGAAGGTGTGGGAGGCCCTTTGAGCCTGGC
CTGCAGAGACCATCCGTGTCCCCTCCCACCTTCATGCCTGTGAGAAGTTAGGAATGTATACGGTACCACATTTGGCA
GTCAGCTTATTTTAATAAATTCAGCAACAGCAAGTCCCTACCATGTTGTGTATCTTCACCATCTTGTCTGACCATGA
CCACTGGCCTTGTGTGTTCTTTTACTCAACGTGTACCCCCGCTCTCCCCCAAAGTGTGGCAGGCTCTCATGCTCCTT
AACCCCCATTGTGGCAATGTCTTACGGGTAACGCTGGAGCTGCAGGAGGAGGGAAGGACACGTCAGAGCCACCAGGC
AGTGGGAGCATCTTGGAGTCCCCACCAGCCTCATGAGGGGGACAGGAAGAGAGCAAATGTGTAGGGAGGAAGGCTGT
GGCTCCTCCCGGGGTGGGAAGGTCAAGCCGATGCTGTCACCCATTTACCAAAGCTGAAGAGAGTGACTTCCTTTCTC
AAAAGCATCACCTTCCCCTGAACCCTGAGTCCAGAGAAGCCAGGAGCCCTCATGTGGCTGCCGAGTTAGCTCAGGGC
TTGGCTTATCACCAACTCTGGTCTCCCTGGGCCAGGGTTGCCAAAACATGAAAGATTTTTTCAGGAGCCAGAGGTTG
GTTCTGACTGGAGGGGGA

<210> SEQ ID NO 411

<211> Length : 117

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 411

>HUMODCA_node_0

GACGTCGGCCCGCCGGCGCCCCACCAGCTCCGCGCGGGCCCGGGTTGGCCACCGCCGGGCCCCCGCCCCTCCCCCGG
CGGTGTCCCGGCCGGAACCGATCGTGGCTGGTTTGAGCTG

<210> SEQ ID NO 412

<211> Length : 110

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 412
>HUMODCA_node_10
ATTGTCACTGCTGTTCCAAGGGCACACGCAGAGGGATTTGGAATTCCTGGAGAGTTGCCTTTGTGAGAAGCTGGAAA
TATTTCTTTCAATTCCATCTCTTAGTTTTCCAT

<210> SEQ ID NO 413

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 413
>HUMODCA_node_12
AGGAACATCAAGAAATCATGAACAACTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTTACTGCC
AAGGACATTCTGGAC

<210> SEQ ID NO 414

<211> Length : 27

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 414
>HUMODCA_node_13
CAGAAAATTAATGAAGTTTCTTCTTCT

<210> SEQ ID NO 415
<211> Length : 72
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 415
>HUMODCA_node_2
CTCCGGCGGGCGGGAGCCAGGCGCTGACGGGCGCGGCGGGGGCGGCCGAGCGCTCCTGCGGCTGCGACTCAG

<210> SEQ ID NO 416
<211> Length : 82
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 416
>HUMODCA_node_27
CTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTTGACA
TGGGG

<210> SEQ ID NO 417
<211> Length : 56
<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 417
>HUMODCA_node_3
GCTCCGGCGTCTGCGCTTCCCCATGGGGCTGGCCTGCGGCGCCTGGGCGCTCTGAG

<210> SEQ ID NO 418

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 418
>HUMODCA_node_30
GCTGAGGTTGGTTTCAGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATT
TGAAGAG

<210> SEQ ID NO 419

<211> Length : 113

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 419
>HUMODCA_node_34
ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTGCATACTC
TATGACCACGCACATGTAAAGCCCCTTCTGCAAAAG

<210> SEQ ID NO 420

<211> Length : 55

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 420
>HUMODCA_node_38
GCAACTCATGCAGCAATTCCAGAACCCCGACTTCCCACCCGAAGTAGAGGAACAG

<210> SEQ ID NO 421

<211> Length : 94

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 421
>HUMODCA_node_40
TTGAAATGTCTTTGTAAGAGTAGGGTCGCCATGATGCAGCCATATGGAAGACTAGGATATGGGTCACACTTATCTGT
GTTCCTATGGAAACTAT

<210> SEQ ID NO 422

<211> Length : 271

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 422
>R00299_node_2
GCGGCCGCAGAGCACTTTGCCCGGAGCCCAGCGTCCTCCCCTGAGTTCGCTGAGTCTCCCGGGACCAGCAAAGGCTG
CGCGCCCCGCATCGGCCCGGAGGCGGGGAGCCCTGGGAGGCCTGGCCGAGCTGCCCGCAGGGAAATGGCGGAGAAAG

CGCTTCTCTGCCCGAGTTCAGCCGGGCTGGGGACTTGGCCCTGGGTCCTGAACTCGGCATGGCCAGTTCTGCCTCTG
GCTGTGGACCAGGGTGTGGACTGGAGACCGCGGGGGCCAG

<210> SEQ ID NO 423

<211> Length : 172

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 423
>R00299_node_30
GGGATCGACATTGAGACCAAGATGCACGTCCGCTTCCTTAACATGGAAACCATGGCCCTCTGCCACTGACCCACCGC
CACCTCCGCGGAGAAACTGCACTTTGCAATGGGGCCGCCTCCCCGCGTAGCTGGAGCAGCCCAGGCCCGGCGGACAG
CCTCTTCCTGCAGCGCCG

<210> SEQ ID NO 424

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 424
>R00299_node_10
GAGAACTTCAACAATGTCCCGGACCTGGAGCTCAACCCCATCCGATCCAAAATTGTTCGTGCCTTCTTCGACAACAG

<210> SEQ ID NO 425

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 425

>R00299_node_14
GAACCTGCGCAAGGGACCCAGTGGCCTGGCTGATGAGATCAATTTCGAGGACTTCCTGACCATCATGTCCTACTTCC
GGCCCATCGACACCACCATGGACGAGGAACAGGTGGAG

<210> SEQ ID NO 426

<211> Length : 25

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 426
>R00299_node_15
CTGTCCCGGAAGGAGAAGCTGAGAT

<210> SEQ ID NO 427

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 427
>R00299_node_20
TTCTGTTCCACATGTACGACTCGGACAGCGACGGCCGCATCACTCTGGAAGAATATCGAAAT

<210> SEQ ID NO 428

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 428
>R00299_node_23
GTGGTCGAGGAGCTGCTGTCGGGAAACCCTCACATCGAGAAGGAGTCCGCTCGCTCCATCGCCGACGGGGCCATGAT
GGAGGCGGCCAGCGTGTGCATGGGGCAGATG

<210> SEQ ID NO 429

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 429
>R00299_node_25
GAGCCTGATCAGGTGTACGAGGGGATCACCTTCGAGGACTTCCTGAAG

<210> SEQ ID NO 430

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 430
>R00299_node_28
ATCTGGCAG

<210> SEQ ID NO 431

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 431
>R00299_node_31
GTACATAGCCAAGGCTCGTCTGCGCACCTTGTGTCTTGTAGGGTATGGTATGTGGGACTTCGCTGTTTTTATCTCCA
ATAAAAAAAAAAAAAAGGTTTGTTAATTAAT

<210> SEQ ID NO 432

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 432
>R00299_node_5
TCTCATCGGATCAGATCGAGCAGCTCCATCGGAGATTTAAGCAGCTGAGTGGAGATCAGCCTACCATTCG

<210> SEQ ID NO 433

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 433
>R00299_node_9
CAAG

<210> SEQ ID NO 434

<211> Length : 157

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 434
>W60282_PEA_1_node_10
GGCTTGTAGGGGGAGAGACCAGGATCATCAAGGGGTTCGAGTGCAAGCCTCACTCCCAGCCCTGGCAGGCAGCCCTG
TTCGAGAAGACGCGGCTACTCTGTGGGGCGACGCTCATCGCCCCCAGATGGCTCCTGACAGCAGCCCACTGCCTCAA
GCC

<210> SEQ ID NO 435
<211> Length : 137
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 435
>W60282_PEA_1_node_18
TACGCCTGCCTCACACCTTGCGATGCGCCAACATCACCATCATTGAGCACCAGAAGTGTGAGAACGCCTACCCCGGC
AACATCACAGACACCATGGTGTGTGCCAGCGTGCAGGAAGGGGGCAAGGACTCCTGCCAG

<210> SEQ ID NO 436
<211> Length : 436
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 436
>W60282_PEA_1_node_22
TCTCTTCAAGGCATTATCTCCTGGGGCCAGGATCCGTGTGCGATCACCCGAAAGCCTGGTGTCTACACGAAAGTCTG
CAAATATGTGGACTGGATCCAGGAGACGATGAAGAACAATTAGACTGGACCCACCCACCACAGCCCATCACCCTCCA
TTTCCACTTGGTGTTTGGTTCCTGTTCACTCTGTTAATAAGAAACCCTAAGCCAAGACCCTCTACGAACATTCTTTG
GGCCTCCTGGACTACAGGAGATGCTGTCACTTAATAATCAACCTGGGGTTCGAAATCAGTGAGACCTGGATTCAAAT
TCTGCCTTGAAATATTGTGACTCTGGGAATGACAACACCTGGTTTGTTCTCTGTTGTATCCCCAGCCCCAAAGACAG
CTCCTGGCCATATATCAAGGTTTCAATAAATATTTGCTAAATGAGTGAATC

<210> SEQ ID NO 437

<211> Length : 669

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 437

>W60282_PEA_1_node_5

GGAGCGGCCTAGGGGAGGCCAGGGGCCCACCTGGGCTGGGGCTGTGGAGAGGGAGTGGCTGGGACGGGAGGAAAAAG
AGAGACGGAGATTAGATGGAAGAAGAGGGATTTCAAGACAAATTGCCAGAGATGCAGTCAGAGAGACTGACTGAGAG
ACACAAAGATAGAAGGAATTAGAGAAAGGGCCACACAGAGCCAGACAGAGAGAGAAGAGTGGAGATGGAGACAGGGA
CGAGGACAGAGAAAGGCAGACAGACACATAGGGACAGAAAGAGAAAAATCACACAAAGTCAGAATTACTGAATGACA
GGGAATGACACATAGAACGAGACACAGATTCAGAGACTCAGGGCAGGGAAAGGAAGGCTGCAGACAGACAGACAGAC
AGAGGGAGGCTGAGACACAGGGAGAAGAGGGGCTTGGAGAGGTGGCACAGGCAGGCAGCCAGTGCCTCAGAGGCCTC
CGGGGAGGGCCCTCACACACACCCCGCCCCGGGGCATTAAGGCAGGGCTTGGAGGCCAGTCATCCTGGGCCCGCCCA
GGGCCGCCCCCCTGCCAGCCCGCCTGCCTGGTGCCTGGCACCTGGCGCTCCAACCCAGCCTACCTGCTGTAGCTGCC
GCCACTGCCGTCTCCGCCGCCACTGGGCCCCCAGAGCCCCAGCCCCAGAGCCT

<210> SEQ ID NO 438

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 438

>W60282_PEA_1_node_21

GGTGACTCCGGGGGCCCTCTGGTCTGTAACCAG

<210> SEQ ID NO 439

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 439
>W60282_PEA_1_node_8
AGGAACCTGGGGCCCGCTCCTCCCCCCTCCAGGCCATGAGGATTCTGCAGTTAATCCTGCTTGCTCTGGCAACAG

<210> SEQ ID NO 440
<211> Length : 616
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 440
>Z41644_PEA_1_node_0
CCTTGCTGTTCATGGCCCAGCAGGGGCCCTATGGGGGTCAGGCCTGCAGGCACTCACACTTGGCACCTGCTCCAAAA
CCCTTTCAGGTCTTTGAGGATCTGAGCCCTGGGCCTGGGTCTCCCGCCGGCTCGGAAAAGCTGGCCTGCCGGGCCAG
ACGAGAGAACCACACGATTCAGAAAAGCAGTGCCCTTCAGCAGCCTCTCCACCGTCTGGGCTCCCCAAAGGCAGAGC
GGGACGCTGGAAATGTGTGCGCGCTGTGGTATGGGTGTGCAAGTGTGCGAAGGCGGCGTGTTGTGTGAGCGAGAGGG
TAGCGGATGTGTGTGTGCGTGTGCGCGCGTGGCTCCGGGTGTGCGCCGCTGCGATAGCGGGTCCTTTCCCGGGGCGG
GCGACGGGCGGGCTGGGAAGGTCTCCTCCCCTCACCACATTGAGAAATCTCAGTGAGTCACCGAGTGGTTCTGCATA
TTAATGAGCTCGCTCGCTGCGAGGGCAGGAGCGGATTTAAAAGAGGCCAGGGCGGGCGGAGGGAGGCTGTGGAGAGA
GCGCGGAGACAAGCGCAGAGCGCAGCGCACGGCCACAGACAGCCCTGGGCATCCACCGACGGCGCAGCCGGAGCCAG

<210> SEQ ID NO 441
<211> Length : 1,062
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 441
>Z41644_PEA_1_node_11
GTACGCCCCGCCTCTACTCACCTTCCTTCCCACCAGACCCAGCTGTGGCTCTCAGGATGGGAAGGGACCCCCCCACC
AGGTCATCTAGCCCCATCTAATATGTGAACACCCACCACAACATCCACAGCAAACAGATACTCAGACAATGCTTACA

TACCCCCAGGGACAAGGAACTCACCACTTACCAAAGCTAGCTATCCATCTCTTGTCCATTTGCAAGCATGGCAGGTT
TGTCATTTTGTAAACTAAAGTCTGTCTCACTCTAATATTTGCATTATAATCTTAATTCCTCTTTTTATTTCAGTTAC
GTAAGTTGTTAAATGGCAGAGTGAGCACTGGCATGGCTGCCAGGGGAGCTCTGAGGACTTCAGTGGGGTGAAATGTG
ACCACTTAGGTGACTGTGTATGTTGGCTATAAAACTGCGCTATAAAACCATGAGGTGCTGAGGATGATCCTTGCCAG
AAACATGTTTTCTTCTCCAAGGTGCCCCACTCCCTCTGCTGCCCAGAAACCTGATAAACTCCTTCCTTCGCAGGTGC
TGGAAGGCACCACAGGTTTGGCTCTTTAAAATCAGAGCCACTGTTAACCAAGGCGGGCAGCAGTGTTAAGACCACCA
GCACCCTGAACCAGCCCTGTACTTACTGGGCACTGTTTCCTTAAAATCAGAAGGTGGCTTCCCATCTCTGGTTTCCT
GGGGTCTTATGTCTGTCCTCGGAGGGAGAATCCAGTTCCTAGCTCCCCTGTACCATGCGAAGGTAGCCTGTCCTGTC
TCACTCCTCAGATACGCAGAGTCTGTTTACACATTTGCCTGCATAGCATGATCAGGAAGCACACACACACACACACA
CACACACACACACGCATGCATGCACACACCATGCAGGTGACTTCCCCAGGAACTAGTGCCAGCACCCCTGCTGCAGA
GGGGGATATCAAGGCTAAATGGAAGAGAGGGGTGACTTGCCTGGGAGCACAGGGCAAAGCCAGGACAGCAAACCAGG
CCTCCTGGTGCTACCCCACCAGCTGCCCTCACAGGGTGGAAGGTACAGCCATAGTGGGTGC

<210> SEQ ID NO 442

<211> Length : 261

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 442

>Z41644_PEA_1_node_12

CTGCATTGCCCTCCCCTCACCTGGCCCAGCCATGCTACCCCAAGCTCAGCCCTGTGACCAGCTCTCCCAGAGCTGAC
ACTCGGGCTCAACCCCTATACCTGAGCCTTTTTTGCTGCCTCCAAAACAGCCTCATCTGCAGTTGCTTGAAATAGAA
AGTGATGAGAGCAATAAATTATTTTCTATAAATCTGCTGGGAATGAAGCCCTCTTTCTGGTCAAGCCAGGCAGCTCA
TGTGGCAAAGGCCAGAACTGCGCAGTCCAC

<210> SEQ ID NO 443

<211> Length : 1,361

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 443

>Z41644_PEA_1_node_15

GCCCTGTGTCATCAAAACTGGCAGAAGGCTAATCCCATGGGCAGGTTATGGAAGAGGCTGAGGGCATCTTGATCTGA
TTGCTGGGGGATACTCAAACCTTTAGCTCACCTTGCTTCTCCCCTCCACCTGAGCTGCAGCCTGGAAAAGGAGGCAC
CCACAGGTCTAAACATGGCCCTGCTTTTTTTTTTCCTGAAAATTCCAATAACAAAAGCAACGAGAGCCTCTCACTAC
CAGGCCTTCTCTCACTTTGCTATAAAATTAGTTCACCCCTCTTTCTTAGAGTGTTGAGGTCCCTGCCCTCCCCACCT
CCCTCCCCTGAAACAAGTTGAAAATATCTTAATGAACATAGAACAGTGATAAAGGAAGTGTTTGAAGTCCTCTTTGT
ACAGAGAGAGAGAGAAAGAGAATGCCAAAGCTAGGTTGGAGGAAGTAGAAGGGTATACGGTGGGCTCAGGCCCATGG
GGGCCACACAGAGGAGCTCTGTGCACTTCAGAGACCAGAGCTTCCAGGGAGCTTCTGGCCACCACAGGAAGCAGCCT
AGTCAGGCATTTTATTTCAATGGATAATTCAGTGGTCTTACTCAGAAATCAAGAACGAGACAGAAAAGTGATAGGCT
AAGTGTAACGTATGGCCCCAGGGCAGCCATGGGGCAGAACTAGAAGAAAGCAAAATATCTAACTGGGCACAGCTTGA
GAGGTGAGGGGAAGGTGGGGCTGGGAACGAGTAGAGATGAGGCAATGCAGCCAGGAGCAGGGACTGAGGGGCACAGG
CCTCCTGCACCACTGCCCCACCCCACCAACCACCTCTTCTGTCTCCAGGAAGCAGCTTCTAGAGCTAGCATTCTTCT
GGAGGACATGCATTATTTGGGCAAAATACAAAGAAATATACAAGCCTAAGTCAAGTAAGGGAATGCCTCCCACCCTT
GCTATTTTCTCTAAATAGAGAGGCTGAGTACAGACGCGGAAAGAAACAAGGAGGTGTGGGAGCAGCCCGCCATGCTA
GAGAAAGACTACATTCCTGCCACTAACAGTCGGTGGCCACTGGGCAAATCTTAAGTCTGTGGTGCCTCAGTTTCCTC
ATATGCAAAGCGGGTTTGTTCCATAGGCCTCTGAGGACAAAATGAGATTGCAGAAGTGAGATTGCAGATGGTTAGAA
AAGACAAAGCCACACTGGTGTGAGTTTTCATGGTCCCCGGGACCACATCCTCAGAAGGATCCCTCCCACTTCTCCTG
GGGGTTCCTGCAGTTCTGGGACAGGGGCATTCCCTGCAGACCAGACGTGAATGAAGCCGCTTAGCCAGCATCTTGTG
AACGGCCTGCCTCATGTCCTGAGCCACTTACACATGTGTTTTTTCTCCCCAG

<210> SEQ ID NO 444

<211> Length : 569

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 444

>Z41644_PEA_1_node_20

ATGGGAGACCCATCTCTCTTGTGCTCCAGACTTCATCACAGGCTGCTTTTTATCAAAAAGGGGAAAACTCATGCCTT
TCCTTTTTAAAAAATGCTTTTTTGTATTTGTCCATACGTCACTATACATCTGAGCTTTATAAGCGCCCGGGAGGAAC
AATGAGCTTGGTGGACACATTTCATTGCAGTGTTGCTCCATTCCTAGCTTGGGAAGCTTCCGCTTAGAGGTCCTGGC
GCCTCGGCACAGCTGCCACGGGCTCTCCTGGGCTTATGGCCGGTCACAGCCTCAGTGTGACTCCACAGTGGCCCCTG
TAGCCGGGCAAGCAGGAGCAGGTCTCTCTGCATCTGTTCTCTGAGGAACTCAAGTTTGGTTGCCAGAAAAATGTGCT
TCATTCCCCCCTGGTTAATTTTTACACACCCTAGGAAACATTTCCAAGATCCTGTGATGGCGAGACAAATGATCCTT

AAAGAAGGTGTGGGGTCTTTCCCAACCTGAGGATTTCTGAAAGGTTCACAGGTTCAATATTTAATGCTTCAGAAGCA
TGTGAGGTTCCCAACACTGTCAGCAAAAAC

<210> SEQ ID NO 445

<211> Length : 163

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 445
>Z41644_PEA_1_node_24
CAATATATTTGTGATTCCCCATGTAATTCTTCAATGTTAAACAGTGCAGTCCTCTTTCGAAAGCTAAGATGACCATG
CGCCCTTTCCTCTGTACATATACCCTTAAGAACGCCCCCTCCACACACTGCCCCCCAGTATATGCCGCATTGTACTG
CTGTGTTAT

<210> SEQ ID NO 446

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 446
>Z41644_PEA_1_node_1
CAGAGCCGGAAGGCGCGCCCCGGGCAGAGAAAGCCGAGCAGAGCTGGGTGGCGTCTCCGGGCCGCCGCTCCGACGGG
CCAG

<210> SEQ ID NO 447

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 447
>Z41644_PEA_1_node_10
CTGCAGAGCACCAAGCGCTTCATCAAGTGGTACAACGCCTGGAACGAGAAGCGCAG

<210> SEQ ID NO 448

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 448
>Z41644_PEA_1_node_13
ACTCTGTCACCCTCCAG

<210> SEQ ID NO 449

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 449
>Z41644_PEA_1_node_16
GGTCTACGAAGAATAGGGTGAAAAACCTCAGAAGGGAAAACTCCAAACCAGTTGGGAGACTTGTGCAAAGGACTTTG
CAGA

<210> SEQ ID NO 450

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 450
>Z41644_PEA_1_node_17
TTAAAAAAAAAAAAAAAAA

<210> SEQ ID NO 451

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 451
>Z41644_PEA_1_node_19
AAGCCTTTCTTTCTCACAGGCATAAGACACAAATTATATATTGTTATGAAGCACTTTTTACCAACGGTCAGTTTTTA
CATTTTATAGCTGCGTGCGAAAGGCTTCCAG

<210> SEQ ID NO 452

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 452
>Z41644_PEA_1_node_2
CGCCCTCCCCATGTCCCTGCTCCCACGCCGCGCCCCTCCG

<210> SEQ ID NO 453

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 453
>Z41644_PEA_1_node_21
CTTAGGAGAAAACTTAAAAATAT

<210> SEQ ID NO 454

<211> Length : 53

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 454
>Z41644_PEA_1_node_22
ATGAATACATGCGCAATACACAGCTACAGACACACATTCTGTTGACAAGGGAA

<210> SEQ ID NO 455

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 455
>Z41644_PEA_1_node_23
AACCTTCAAAGCATGTTTCTTTCCCTCACCACAACAGAACATGCAGTACTAAAG

<210> SEQ ID NO 456

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 456
>Z41644_PEA_1_node_25
ATGCTATGTACATGTCAGAAACCATTAGCATTGCATGCAGGTTTCATATTCTTTCTAAGATGGAAAGTAATAAAATA
TATTTGAAATGTACCAAAATTCTAGA

<210> SEQ ID NO 457
<211> Length : 36
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 457
>Z41644_PEA_1_node_3
GTCAGCATGAGGCTCCTGGCGGCCGCGCTGCTCCTG

<210> SEQ ID NO 458
<211> Length : 34
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 458
>Z41644_PEA_1_node_4
CTGCTGCTGGCGCTGTACACCGCGCGTGTGGACG

<210> SEQ ID NO 459
<211> Length : 106
<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 459
>Z41644_PEA_1_node_6
GGTCCAAATGCAAGTGCTCCCGGAAGGGACCCAAGATCCGCTACAGCGACGTGAAGAAGCTGGAAATGAAGCCAAAG
TACCCGCACTGCGAGGAGAAGATGGTTAT

<210> SEQ ID NO 460

<211> Length : 58

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 460
>Z41644_PEA_1_node_9
CATCACCACCAAGAGCGTGTCCAGGTACCGAGGTCAGGAGCACTGCCTGCACCCCAAG

Segment nucleic acid sequences:

<210> SEQ ID NO 461

<211> Length : 669

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 461
>Z44808_PEA_1_node_0
CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG

GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACG

<210> SEQ ID NO 462

<211> Length : 187

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 462
>Z44808_PEA_1_node_16
TGTCATCCTGTGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCT
GAGTGTGCGCACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGA
CACGGGGCGCCCCATTCCCGGCACATCCACAAG

<210> SEQ ID NO 463

<211> Length : 172

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 463
>Z44808_PEA_1_node_2
TTTTTGAGAGTGGATCAAGATAAAGACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGC
ATCTGACGGAAGGACCTTCCTTTCCCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCAT
ATCGAGGAAACTGCAAAG

<210> SEQ ID NO 464

<211> Length : 275

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 464

>Z44808_PEA_1_node_24

GCTCTCAGAACCCGACCCCAGCCATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACT
CCAGTGGAGACATCGGCAAAAAGGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGT
GTGAAGAAGTTTGTTGAATACTGTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGG
CGTGGCGAAAGAGGACGGCAAAGCGGACACCAAGAAACGCCACA

<210> SEQ ID NO 465

<211> Length : 1,685

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 465

>Z44808_PEA_1_node_32

GATGCAATGGTGGTGTCCTCCAGACCCAAAGCCACAACCCATCGCAAGTCAAGAACACTTTCCAGAAGATAAACATG
AGTGGGTTCATGTCTCTCTCCTTCAAAGCCAGGACAAAATCCCCACTTCTTTGCTGCCGCGAGTCAATTTGTGATTT
ATTTTGTCTGCACCTGTTTGATGCCAGGTCGACATTTCCTAAGGCAAGCCCCTGTATTTGTTGTGGATTTAAGTGGA
GGCGGCCAGCACACACCTTGGATGTAATTTAAAACCATTTCCTGAGGAAAGATGTGTGATATGCTTTCCTTTGTTTA
GCAAATGTTTATGGTTTTAACTTTAAATCTCACCGCAAATCACTTACACTTGAAAACAGGGCTGGTCTGAAAGTAAT
TACCCTCCCTGAGTGCCAAGACCTCCAGAAGTTGTTTTCATTCCCGAATGGCAATCACTGTACTCATGCGCTCCACG
CATCTTAAATAAACTCAGTTCAAAGCACATGCCTCCTGCTTCAGCTCTTTTTTCCAAAAAGAGAAACAGAAGCAGGT
TCCCCCTCCTTTTATAGTGCCTGCGTGGACACGCGGACCTCCATGCCTTTCATGCTGTGGCTATGTCAGCAAACTAC
GATATTGGGATGATCCTAACGGGCAAGCCAGCTGCGGCTCCTACCGGCCGTGGCCATTGAAGGCCACCATGTTGCTT
TGAAACATCTCAAAGAATAACATAGTGCCAGCCAGCAAGGGTTTCACCATATGCATGACCCAGACAGGAACTATCAA
AAGAAGGGATCACGGGAAGGTGCATGATGCTAATGTGGAATCCAGAGGAGCTCTTTCCTGATCTCTTCAGCTTCCGC
TGCCACTCCAGAATCATCAGAGCTGATATTAAATAAGTTAAAATGTTAGTCCACCGTCTCCTCCTGCAATCCTAACC
ATCTTTTGAGACTGTTAGAATACTTTGACGGGTTGTCTTTCTGTGCAACTAATTTAAACCTCAAGTTTAGTGTAGGA
GATGGGTTTGTCTTCTCACCTCTTCAGATCTTTATCAAGGGGGAATAAAAGCCAACCCAGAAACCTAAACTTTAAAA
TTTAATTATTTGAAATAATAAAACAGAAGAAGGGATCAACATTTGTCGGAATTGGCACTCTTGGAAAACTAAGTCTA
GGAGATCATATATTGCTTTTTTTTTTTCATTCTAAATTACTTTTAATTGAAAGTCAAGATGCTGAGTTACAGTTGTT

TATCATTATAATAAGCAAACTTTTTAAGTTGGATTTCTTCTTAAAGAGGTAAACTAGTGAACAAAAAAGATAAAAAG
GAAAATTAAGAATCAACTATGCCTTTATCAAATTTGAAGCATAAGTTATATTATTAAAATTATTTTTGTATAATCAA
GGTGATAAGACATTCTGGAAAACATTTAATGTATTTAGTACTTAGAATATTTACAGTGGATGTTACTTTTTTGAAAC
GATATATTTTTCCCAATTTTTCTATCATGTCAAGGAAGGAAACTGTTAAGAAGTTACCAGTGTCCAAAATGTCTTCA
TTGTTTCTTACTCATACTTACACCTCACATGACCTGCCCAGCCCTCTTTGGTTCAGTTCATTCCCAGAAGCCAAGCC
TTAGTCTTCACAGATGAGCGACACACACCTCTGAATATAATGTCTCTTTTTTGTTTTTTCCTTTTCAG

<210> SEQ ID NO 466

<211> Length : 877

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 466

>Z44808_PEA_1_node_33

CCAAGGAAACAAGGATAAATGGCTCATACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCACCAAAGAGC
AATTAAGAAAACAAAAACAGAAACACATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGAG
CTATTTAAACAGACTGTTTTAATCTGTGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAATGTATGTGTC
CTCTTTTGACCTTGGAAATCTGTATGTGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCGCCTTCCATA
TGTTAACAGTAGAGCTCTATGCACTCCGGCTGCAATCGTATGGCTTTCTCTAACCCCTGCAGTCACTTCCAGATGCC
TGTGCTTACAGCATTGTGGAATCATGTTGGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCCGACCCTACA
GGCAGTTAACATGCATGGGCTGGTTTGTTTCTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTTCCAGAGATCTT
GCTCATACAATGAATCACGCAACCACTAAAGCTATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAATAATATTTT
CAAACTGTCGTTGGTGTGATACTTTGGCTCAAAGGATCTTTGCTTTTCCATTTTAAGCTTCTGTTTTGAGTTTTGCC
CTGGGGCTTGAATGAGTCCCAGAGAGTCGTTCGGATGGTGGGAGGCTGCCTAGGAGGCAGTAAATCCAGTCACAGTG
CCTGGGAGGGGCCCATCCTTCCAAAATGTAAATCCAGTCGCGGTGTGACCGAGCTGGCTAACAGGCTTGTCTGCCTG
GTTTTCCTCCTACACGTGGACATTATTCTC

<210> SEQ ID NO 467

<211> Length : 252

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 467
>Z44808_PEA_1_node_36
GATACGTGGTGCCCCCGGGGCTGGTGTTGGCAGCCGGGGGGAGGTGCCTGAGGGTCCCCACGGTTCCTTTCTGCTTT
TCTGAATGCATCAAGGGTACGAGAACTTGCCAATGGGAAATTCATCCGAGTGGCACTGGCAGAGAAGGATAGGAGTG
GAATGCCCACACAGTGACCAACAGAACTGGTCTGCGTGCATAACCAGCTGCCACCCTCAGGCCTGGGCCCCAGAGCT
CAGGGCACCCAGTGTCTTAAG

<210> SEQ ID NO 468

<211> Length : 349

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 468
>Z44808_PEA_1_node_37
GAACCATTTGGAGGACAGTCTGAGAGCAGGAACTTCAAGCTGTGATTCTATCTCGGCTCAGACTTTTGGTTGGAAAA
AGATCTTCATGGCCCCAAATCCCCTGAGACATGCCTTGTAGAATGATTTTGTGATGTTGTGATGCTTGTGGAGCATC
GCGTAAGGCTTCTTGCTTATTTAAACTGTGCAAGGTAAAAATCAAGCCTTTGGAGCCACAGAACCAGCTCAAGTACA
TGCCAATGTTGTTTAAGAAACAGTTATGATCCTAAACTTTTTGGATAATCTTTTATATTTCTGACCTTTGAATTTAA
TCATTGTTCTTAGATTAAAATAAAATATGCTATTGAAACTA

<210> SEQ ID NO 469

<211> Length : 0

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 469
>Z44808_PEA_1_node_41
CACCTCATAATCATGTGAAAAAGACACTCAAAAACTACCATTTGAATGGATGGATGAAAATAACCTCCGTATATTCT
ACGAAGATGTTTAATAATAAATAGGTTTCGTTATAAGAGAATGTGTGTCACTTCGTCTCTTCCCTCACCCCCGAGAC

TTAGTGACAGTTATTTTTGACTTTTCCAACTATACTATTTGCCTAGAAAATGTGTCTATTAAATAGCGTATTGAGAA
AT

<210> SEQ ID NO 470

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 470
>Z44808_PEA_1_node_11
ATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAG

<210> SEQ ID NO 471

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 471
>Z44808_PEA_1_node_13
ATATTGCATCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAG

<210> SEQ ID NO 472

<211> Length : 83

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 472

>Z44808_PEA_1_node_18
GTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCGGGACCTGTACAAGGGCCGCCAGC
TACAAG

<210> SEQ ID NO 473

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 473
>Z44808_PEA_1_node_22
GTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACGCGCTGTCCACGGACATGGTCCACGCCGCC
TCCGACCCCTCCTCCTCGTCAGGCAG

<210> SEQ ID NO 474

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 474
>Z44808_PEA_1_node_26
GAAGTAAGAGAAACCTGTGATGGCCAGAGCCCAGATGTTCTTAGGAGGCAAGCCAGGAGAAGCCGGGTCTGACTTTT
CAGCTCAGAGACAGCACT

<210> SEQ ID NO 475

<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 475
>Z44808_PEA_1_node_30
CCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACAG

<210> SEQ ID NO 476

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 476
>Z44808_PEA_1_node_34
CTGATCCTCCTACCTGGTCCACCCCAGGGCTACCGGAAGGTAAAATCTTCACCTGAACCAATTATGAGCAGTCTC

<210> SEQ ID NO 477

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 477
>Z44808_PEA_1_node_35
CTTACTGAAGGTACAGCCG

<210> SEQ ID NO 478

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 478
>Z44808_PEA_1_node_39
CTACTGTGGTTGAGAGGAAAGGTGTCTTTTTATTGCTTCTAGAGACGTTGAAAGTGTGACCTGAG

<210> SEQ ID NO 479

<211> Length : 107

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 479
>Z44808_PEA_1_node_4
ACGTGTCCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCT
GAGTGCAATGACGACGGCACCTACAGTCAG

<210> SEQ ID NO 480

<211> Length : 100

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 480
>Z44808_PEA_1_node_6
GTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAGGCCCATCAGCGGCACTGCCGTGGC
CCACAAGACGCCCCGGTGCCCGG

<210> SEQ ID NO 481

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 481

>Z44808_PEA_1_node_8

GTTCCGTAAATGAAAAGTTACCCCAACGCGAAGGCACAGGAAAAACAG

Segment nucleic acid sequences:

<210> SEQ ID NO 482

<211> Length : 170

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 482

>AA161187_node_0

GCTGGGAGTAGAGGGCAGAGCTCCCACCCCGCCCCGCCCCCAGGGGGCGCCCCGGGCCCGGCGCGTTAGGAGGCAGA

GGGGGCGTCAGGCCGCGGGAGAGGAGGCCATGGGCGCGCGCGGGGCGCTGCTGCTGGCGCTGCTGCTGGCTCGGGCT

GGACTCAGGAAGCCGG

<210> SEQ ID NO 483

<211> Length : 120

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 483

>AA161187_node_6

GCACACACGCGAGGGGACCCTGGGTGGGCAAAAACGTGCTTTCCCGGACGGGGTTGAAGGGGAGAAAGGGAGAGGTC

GGGCTTGGGGGGCTGCCTCCCGCGGCTCAGCAGTTCCTCTGAC

<210> SEQ ID NO 484

<211> Length : 211

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 484
>AA161187_node_14
GCCTACTACACCCGTTACTTCGTATCGAATATCTATCTGAGCCCTCGCTACCTGGGGAATTCACCCTATGACATTGC
CTTGGTGAAGCTGTCTGCACCTGTCACCTACACTAAACACATCCAGCCCATCTGTCTCCAGGCCTCCACATTTGAGT
TTGAGAACCGGACAGACTGCTGGGTGACTGGCTGGGGGTACATCAAAGAGGATGAGG

<210> SEQ ID NO 485

<211> Length : 297

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 485
>AA161187_node_16
TGCTCACCAATGCCCCAGGCATCAGGCTCCTGGGCTGCCTCTCCATGCCTCCCACACCCACCCTAGCTCTGGCCGAT
TCTCCTGCAGCACTGAGCCCATTCCTCTCCCCAGAAACTTCCAAGCCATGCTCAACCGCAGCTCCCACGGAAACCCC
TCTGGGGGTTCCTCTGGTGGGCCTGCCCTGGCACCTGCGTGTCCCCCAACACACATGCCCTGAAAGAAGTGGGCCCA
GCATCCGGAGGAGCCCCGGCAGCCCCAGACTGGGCGTGTTCCCTGTATCAGGAATCCCTTCCCTCT

<210> SEQ ID NO 486

<211> Length : 225

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 486

>AA161187_node_25

GAAAGCATCCTGTGTCCCTGTGCCTTATTTGACCCTCATGCCAACCCCGGGAGGTGGAGACTGTTGCCCCACTCTGC
AGATGCAGAAACGGAGGCTTGGCTGCTGCCAGGGGGAGGAGGAGGATGTGCACCCAGTCTACCCAGCCCCATAGCCC
TTCCCACTCTCAGCCCCTCCCCTGCCCCACTCACTCTGCCCCAGGCTGACCTCAGCCCCGCTGCTCCCCAG

<210> SEQ ID NO 487

<211> Length : 362

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 487

>AA161187_node_26

GGTGACTCAGGTGGACCCTTGGCCTGTAACAAGAATGGACTGTGGTATCAGATTGGAGTCGTGAGCTGGGGAGTGGG
CTGTGGTCGGCCCAATCGGCCCGGTGTCTACACCAATATCAGCCACCACTTTGAGTGGATCCAGAAGCTGATGGCCC
AGAGTGGCATGTCCCAGCCAGACCCCTCCTGGCCACTACTCTTTTTCCCTCTTCTCTGGGCTCTCCCACTCCTGGGG
CCGGTCTGAGCCTACCTGAGCCCATGCAGCCTGGGGCCACTGCCAAGTCAGGCCCTGGTTCTCTTCTGTCTTGTTTG
GTAATAAACACATTCCAGTTGATGCCTTGCAGGGCATTCTTCAAAAGCAATGGC

<210> SEQ ID NO 488

<211> Length : 515

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 488

>AA161187_node_28

ATAAGAGGACACAGTGAGAAGATGGGGGTCTGCTCGCCAGGAAGGAGCCCTCACCAGCAGCCGCATCGCTCAGCACC
TTGATCCTGGACTTCCAGCCTCCAGAGCTGTGAGAAACAAACCTCTATCATCTACCAGCCGCCCACGGCGTGGGATT
TGTGTTACAGCAGCCTGAGCTGACCCAGACGCCAAGGAGCAACACACGCACCAGGGTAGGCTGGAGAAACCAGAACC
CGGGAATCCCGCCTCCCTCAACTTGAAACTTGGGAATAGTGTATTCTCTTTTCAACACTTGCACTAGTAGAAGGTTA
ATTACATGAAAGATTAGGCAAAATGTATGGCTATGTGTCCTGGTTTTCCAATAAAAGTATTGAGTTTCTCTGGGGAA

AGTGCAGATAAAATGCTTAGTGGAGGCTGGGCGCTGTGGCTTATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGC
AGGCAGGCAGATCACAAGGTCAGGAGTTTGAGACCGGCCTGGCCAATATGATG

<210> SEQ ID NO 489

<211> Length : 27

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 489
>AA161187_node_4
AGTCGCAGGAGGCGGCGCCCTTATCAG

<210> SEQ ID NO 490

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 490
>AA161187_node_7
CATCCGAG

<210> SEQ ID NO 491

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 491

>AA161187_node_8
GACCATGCGGCCGACGGGTCATCACGTCGCGCATCGTGGGTGGAGAGGACGCCGAACTC

<210> SEQ ID NO 492

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 492
>AA161187_node_9
GGGCGTTGGCCGTGGCAGGGGAGCCTGCGCCTGTGGGATTCC

<210> SEQ ID NO 493

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 493
>AA161187_node_10
CACGTATGCGGAGTGAGCCTGCTCAGCCACCGCTGGGCACTCACGGCGGCGCACTGCTTTGAAAC

<210> SEQ ID NO 494

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 494
>AA161187_node_12
CTATAG

<210> SEQ ID NO 495
<211> Length : 76
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 495
>AA161187_node_13
TGACCTTAGTGATCCCTCCGGGTGGATGGTCCAGTTTGGCCAGCTGACTTCCATGCCATCCTTCTGGAGCCTGCAG

<210> SEQ ID NO 496
<211> Length : 37
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 496
>AA161187_node_19
GTTGCTGTCTCTCTCCTTCCCACTATCGTCCGCACAG

<210> SEQ ID NO 497
<211> Length : 26
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 497
>AA161187_node_20
CACTGCCATCTCCCCACACCCTCCAG

<210> SEQ ID NO 498

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 498
>AA161187_node_21
GAAGTTCAGGTCGCCATCATAAACAACTCTATGTGCAACCACCTCTTCCTCAAGTACAG

<210> SEQ ID NO 499

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 499
>AA161187_node_22
TTTCCGCAAGGACATCTTTGGAGACATG

<210> SEQ ID NO 500

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 500
>AA161187_node_23
GTTTGTGCTGGCAATGCCCAAGGCGGGAAGGATGCCTGCTTC

<210> SEQ ID NO 501
<211> Length : 31
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 501
>AA161187_node_24
GTGAGTGTCCCTGCCACCACTCCCAGCCCAG

Segment nucleic acid sequences:

<210> SEQ ID NO 502
<211> Length : 712
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 502
>R66178_node_0
GCAGCGAGCGCGGCTCCACATTGTTGCGGATCGCCGGCACCCGGCAGAGCGGCGGCGGCTGGGACGCGCGGCGCCTC
CGACCCGTTCTCCTCGCGCCCGGCCGCGCAGCCAGAGCCACCCGGGCCGCCGACCGCCGAGCCCCGCGCCCGCCGCC
TGGGCCCCGAGCCTTCTGCGCCGCCCGGGTGCGTCCCGCCACCCTCGGAGGACGGCCGGCCATGGACGCCTGCAAGT
TGGAGCCGAGCGGGAGGGTGTGAGCGGGCCGGGGCCAGGAGCCCGCGCCGCGCAACCGGGCAGCCGGGCGCGCCGGG
GGTGGGTCCCTCTCCCCAGCCCCGCTCTGCGTGGAAGAAGAGGGCGGGGACCGGCGCCGGGAGGAGAGCGGAGGAGG
CGAAGGGGCATGACTCGTGCAACTTGCGGCGGGCATCTGCCGAGCCTCTGAGCCGGCGGCGGCCCGGGGCCCGGACT
GCGGCCGCGCGGATCCACCCAGCCCACCCCGCCCCGGCCGACGGCTGCAGCTGACCTGGATCCTTCGAGCGCCCGCC

GACCGCCAGCGATCTTCCCTCATCTTCCGGGCTGGTTTCTGCTGCGCGAGGAGCGCTGCCCTCGCCGCCCCTCTCGC
CGGACCCCCGGCCCCCGATGGCTCGGATGGGGCTTGCGGGCGCCGCTGGACGCTGGTGGGGACTCGCTCTCGGCTTG
ACCGCATTCTTCCTCCCAG

<210> SEQ ID NO 503

<211> Length : 302

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 503
>R66178_node_6
CGGCACAGACGTGGTTCTGCACTGCAGCTTTGCCAACCCGCTTCCCAGCGTGAAGATCACCCAGGTCACATGGCAGA
AGTCCACCAATGGCTCCAAGCAGAACGTGGCCATCTACAACCCATCCATGGGCGTGTCCGTGCTGGCTCCCTACCGC
GAGCGTGTGGAATTCCTGCGGCCCTCCTTCACCGATGGCACTATCCGCCTCTCCCGCCTGGAGCTGGAGGATGAGGG
TGTCTACATCTGCGAGTTTGCTACCTTCCCTACGGGCAATCGAGAAAGCCAGCTCAATCTCACGGTGATGG

<210> SEQ ID NO 504

<211> Length : 206

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 504
>R66178_node_8
CCAAACCCACCAATTGGATAGAGGGTACCCAGGCAGTGCTTCGAGCCAAGAAGGGGCAGGATGACAAGGTCCTGGTG
GCCACCTGCACCTCAGCCAATGGGAAGCCTCCCAGTGTGGTATCCTGGGAAACTCGGTTAAAAGGTGAGGCAGAGTA
CCAGGAGATCCGGAACCCCAATGGCACAGTGACGGTCATCAGCCGCTACCGC

<210> SEQ ID NO 505

<211> Length : 139

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 505

>R66178_node_15

GCTAAATGGCTCTCTCCCCAAGGGTGTGGAGGCCCAGAACAGAACCCTCTTCTTCAAGGGACCCATCAACTACAGCC

TGGCAGGGACCTACATCTGTGAGGCCACCAACCCCATCGGTACACGCTCAGGCCAGGTGGAG

<210> SEQ ID NO 506

<211> Length : 1,474

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 506

>R66178_node_24

GTGAGGGTCACAGTCTGCCCATCAGTCCTGGAGTTCTTCAGACCCAGAATTGCGGGCCTTGAAGGCCAGTGTCCTGG

AAGGGAGGCAGGTGTGAGCCTGCAAGTGTGCATGCCCAGCCATGGTATGGACATGTGTCTCTGGGCATGTAAATGTG

AACCAGTGTGAACAGGCTCCGTCTGCATGCTGAGTGTGCATGTGGGAGCCCGTGGCTGTGCCGTGGCAACGTGCCAT

TCTCTGAGCCAGCGAATGGCAGTGTGTTGGGAGGTCTGAGAAGGCAGCTGCATCCGTGCCTCTGGGAGGATTCGGTT

CTCCCCCAGCTTGCCGAGGCCCTGCCTGATGGTCTGACACGAGGCACAGCTGCTGCAGCTGCAGATGGACAGAAGGG

CTTCCCAGAGGTGGACCCCAGGCCTCCCCACTCTCCCTGTGGCTGGCTGCACTGCATGCTGGGGGGTGTAGTTCTTG

CAGCTTCCAGGCCTAATCTGATGCCGGAGCATTTCCTGCCTGAGGAAGCGCCAGGCATTGGTTTCGGAGGCAAACCC

AAACATTCTCTTTGACCCCAAACCTCCAGATCCTAGATCCAGACTGTAAGCCCTAACACTTCACTCCACCTCAGATC

TATCCAAAGCCCCCAGCACCAGCCCACCCACCTCAGTCAGAGAACCAGGACCCCAAAGGCATGCAGAGCCCCCACTT

CCCCACTGTCTTGGCCAGCCAGGGACCCCAGAAGAGAGGTTACAACCCTTCAGGAATAGGGACAAGCTGCTCCCTTT

GTAAGAGGATGTGAGGGAGGCTGGCTGGGCCCCTGCCAGCAAACACAAATGACCCTGCGGCCTGGCTCTTCTCTCTC

CTCCCAGCTGCGGCCCTTGCAGCTCTGCTCCTGGCACAGAGACAGGAGCTACTGGCTGAGTGTAACAGCTGGAGGGA

TGGAGGGGGAGGGGAGGACGCTCCACTCCACGCCAGACAGCCCCTTCTGCTTGCAAATGAGTTAGATCCCCATGCTT

CTCCTTTCTTCTCTCCCTCACTCAGTATCCTCACTCGAAAGTCTTTGATGCTGGAAGGTCATCCCAGCCATTCTGCT

GCTGCTACACAGGCCCAGCCCTAAACAAAATAACCGGGGTTCTTTGGTCCCAAAAGATCCCCAGGAAAGAGTAAACC

TCCTTAAGACTTAAGGAAAAAAGTTGGCTAGGCGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGG

TGGGCAGACCACTTGAGATCAGGAGTTCGAGACCAGCCTGGCCAACGTTGTGAAACCCCGTCTCTACAAAATATTAG

CCGGGTGTGGCAGTGCGTGCCTGTAGTCCCAACTACTCAGGAGGCTGAGGCACAAGAATTGCTTGAGCCTGGGAGGC

GGAGGTTGCAGTGAGCCGAGATCGTGCCACTGCACTCCAGCCCAGGCGATGGAGCGAGACTCTGTCCCGCAAAAAAA
AAAAAAAAAAA

<210> SEQ ID NO 507

<211> Length : 464

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 507
>R66178_node_26
AATTCCCCTACACCCCGTCTCCTCCCGAACATGGGCGGCGCGCCGGGCCGGTGCCCACGGCCATCATTGGGGGCGTG
GCGGGGAGCATCCTGCTGGTGTTGATTGTGGTCGGCGGGATCGTGGTCGCCCTGCGTCGGCGCCGGCACACCTTCAA
GGGTGACTACAGCACCAAGAAGCACGTGTATGGCAACGGCTACAGCAAGGCAGGCATCCCCCAGCACCACCCACCAA
TGGCACAGAACCTGCAGTACCCCGACGACTCAGACGACGAGAAGAAGGCCGGCCCACTGGGTGGAAGCAGCTATGAG
GAGGAGGAGGAGGAGGAGGAGGGCGGTGGAGGGGGCGAGCGCAAGGTGGGCGGCCCCCACCCCAAATATGACGAGGA
CGCCAAGCGGCCCTACTTCACCGTGGATGAGGCCGAGGCCCGTCAGGACGGCTACGGGGACCGGACTCTGGGCTACC
AG

<210> SEQ ID NO 508

<211> Length : 277

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 508
>R66178_node_27
TACGACCCTGAGCAGCTGGACTTGGCTGAGAACATGGTTTCTCAGAACGACGGGTCTTTCATTTCCAAGAAGGAGTG
GTACGTGTAGCCCCCCTTCCAGAGCCTCTGTCTGTGACCGCTCCTAACCAGCCCCTCCCCGCACGCCCCCTGCCCAC
CCCCCACCTCCCACTCCAGGAGCTGAACAGAGACTTGCCCAGCTGCCCAAAGCCAGCCCCGAACTCCTGGGGGGCCA
GGGGAGCCCAGGGCAGCCACGACTTGGCTTTGTGTTTTATTTCCTC

<210> SEQ ID NO 509

<211> Length : 37

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 509
>R66178_node_4
GCGTCCACTCCCAGGTGGTCCAGGTGAACGACTCCAT

<210> SEQ ID NO 510

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 510
>R66178_node_5
GTATGGCTTCAT

<210> SEQ ID NO 511

<211> Length : 97

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 511
>R66178_node_9

CTGGTGCCCAGCAGGGAAGCCCACCAGCAGTCCTTGGCCTGCATCGTCAACTACCACATGGACCGCTTCAAGGAAAG
CCTCACTCTCAACGTGCAGT

<210> SEQ ID NO 512

<211> Length : 118

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 512
>R66178_node_11
ATGAGCCTGAGGTAACCATTGAGGGGTTTGATGGCAACTGGTACCTGCAGCGGATGGACGTGAAGCTCACCTGCAAA
GCTGATGCTAACCCCCCAGCCACTGAGTACCACTGGACCAC

<210> SEQ ID NO 513

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 513
>R66178_node_16
GTCAATATCACAG

<210> SEQ ID NO 514

<211> Length : 107

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 514
>R66178_node_18
CTTTCTGTCAACTTATCTATCCGGGCAAAGGGAGGACAAGAGCTAGGATGTTCTGAGGAGAGACTTCACCTGGGACG
TGAAAGGAGCATGGGCTTGATGTCAGACAG

<210> SEQ ID NO 515

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 515
>R66178_node_19
CTGTGACCCTGGACAGGGCC

<210> SEQ ID NO 516

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 516
>R66178_node_20
CCCCCCACCATCTGTAAAACGGGGACAG

<210> SEQ ID NO 517

<211> Length : 112

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 517

>R66178_node_21

TATGATGTACCTTGAAGGGCTGTTGTCAGAATTCTACGTGATGTAAGTCAAGCACCTAGCACAGATCAGTCTGTCAA

TAAATGGCCAATGTTCCGTGATTATTACCCTGACC

Segment nucleic acid sequences:

<210> SEQ ID NO 518

<211> Length : 264

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 518

>HUMPHOSLIP_PEA_2_node_0

GGGTCTCCCACTTGTCCAGACAGCGGCCGGGCTTGTCACGGGGCTCTGTGCAGCCTTTTCCACTCTCCCGGCTGCCA

GCGTCCCGCCCCGTCCCCTCCCAGCCCCCAAGGGAGGAGGGGAGAGCTGCAGAGAGGAGGAGGGGTCGGGGAGGCCC

GGCTTTATAAAGGCGGCTGGAACAACCCTGCCCGCCAGACCCCGTCGCCCGGATCCCCTGAGCTGCCCGCCATCCCA

CGTGACCGCGCCGCCCCCCAGCTCCACCGCTGA

<210> SEQ ID NO 519

<211> Length : 156

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 519

>HUMPHOSLIP_PEA_2_node_19

CTATGATGGGGGCTACATCAACGCCTCAGCTGAGGGTGTGTCCATCCGCACTGGTCTGGAGCTCTCCCGGGATCCCG
CTGGACGGATGAAAGTGTCCAATGTCTCCTGCCAGGCCTCTGTCTCCAGAATGCACGCGGCCTTCGGGGGAACCTTC
AA

<210> SEQ ID NO 520

<211> Length : 141

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 520
>HUMPHOSLIP_PEA_2_node_34
CTCCCCAACCGGGCAGTGGAGCCCCAGCTGCAGGAGGAAGAGCGGATGGTGTATGTGGCCTTCTCTGAGTTCTTCTT
CGACTCTGCCATGGAGAGCTACTTCCGGGCGGGGGCCCTGCAGCTGTTGCTGGTGGGGGACAAG

<210> SEQ ID NO 521

<211> Length : 419

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 521
>HUMPHOSLIP_PEA_2_node_68
CTATCAATTACAGTCCGTCCACCACCTCCCTGTGGGCTGTCCTGAGCTCTGTTGGGTTCCTGGGATGGAATCAGTGC
ATCATAAAGGGCATTCTTTAAGCAGAGAAGGGGCCAGGCCACCCCATTCAGGAACTGCTGCGGGAATAAAGTGCTAA
CTTGCCCCCAGGCTGTCTATGGGAGACCCTGGGCCCAGTCTGGGATGTACAGGGCTCTGGGAAGGGGGCAGTCCTGG
CGGCAGAACCCGGCCTGCAGGGGCACTTTGCTTAGAAGAGGACTCTCCTAGCGGGAGAGGCTGGGAGGGGCTGCATC
AGGCCGTGGAGCTGGTTGCTGTGGTCATCAGTATGGCTGCTTGTTCAGGAAGCGGGAGAACATGGTGAAGGCAGCGA
GGGGCTTGTCGGTGGGAACCATGTGGCCGGCGCC

<210> SEQ ID NO 522

<211> Length : 232

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 522
>HUMPHOSLIP_PEA_2_node_70
GTTCGTGAGCTCCTGACCCCACCATTCCCTCCTCCCCATATAACTGCTCACTCGGGGGCAATTCCTTCATCCCAAAC
CCTTTATTCTTCCCAGAACCCTCCCCACCCCTCTCCAAAAAAACTTGCCCATACAGGGGCCAGATGGTGACCCATGA
CCCAGCCTAAAAGGCAGCCAGAGGGAAAGGACGGGTGGGTCCTGCTCCTTTGCCTCCGGCCCAGTTATCTCTCAGCA
G

<210> SEQ ID NO 523

<211> Length : 280

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 523
>HUMPHOSLIP_PEA_2_node_75
CTTCTGGTTGAGGGAATCCACAAACCACTCATCCCCCATGAAATTGCAGGCCATGTCTACATCTCCATTATATAATA
GGATCTGGTATTTCTAAAGCAGGATGGGGTAAAAATGAGGGGTGTGGAACAAGCCCAGTCCCCAGCCCTTCCCTAGT
TCAAGGCCTACCCCTCAGGAAATTCAAGGGGCCAAGCTAGATAACACGAACCAGGGAATTTTCATGTTTTCTAACGA
CTTACTGCATGTCCAGTATTCTACTAAATGTTTTATCTGTGAAAGTAGA

<210> SEQ ID NO 524

<211> Length : 73

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 524
>HUMPHOSLIP_PEA_2_node_2
GCCCGCTCGCCATGGCCCTCTTCGGGGCCCTCTTCCTAGCGCTGCTGGCAGGCGCACATGCAGAGTTCCCAGG

<210> SEQ ID NO 525
<211> Length : 18
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 525
>HUMPHOSLIP_PEA_2_node_3
CTGCAAGATCCGCGTCAC

<210> SEQ ID NO 526
<211> Length : 20
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 526
>HUMPHOSLIP_PEA_2_node_4
CTCCAAGGCGCTGGAGCTGG

<210> SEQ ID NO 527
<211> Length : 8
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 527
>HUMPHOSLIP_PEA_2_node_6
TGAAGCAG

<210> SEQ ID NO 528

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 528
>HUMPHOSLIP_PEA_2_node_7
GAGGGG

<210> SEQ ID NO 529

<211> Length : 35

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 529
>HUMPHOSLIP_PEA_2_node_8
CTGCGCTTTCTGGAGCAAGAGCTGGAGACTATCAC

<210> SEQ ID NO 530

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 530
>HUMPHOSLIP_PEA_2_node_9
CATTCCGGACCTGCGGGGCAAAGAAGGCCACTTCTACTACAACATCTCTGA

<210> SEQ ID NO 531

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 531
>HUMPHOSLIP_PEA_2_node_14
GCCTGGACTTGAAAGGGGAGCAGACAAATTTCCTGTCGTTGGGGGAAGTTCCCTCTTCTTGGCCCTGGATCTGACCC
TGAGGCCTCCTGTAG

<210> SEQ ID NO 532

<211> Length : 16

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 532
>HUMPHOSLIP_PEA_2_node_15
GGTGAAGGTCACAGAG

<210> SEQ ID NO 533

<211> Length : 89

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 533
>HUMPHOSLIP_PEA_2_node_16
CTGCAACTGACATCTTCCGAGCTCGATTTCCAGCCACAGCAGGAGCTGATGCTTCAAATCACCAATGCCTCCTTGGG
GCTGCGCTTCCG

<210> SEQ ID NO 534

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 534
>HUMPHOSLIP_PEA_2_node_17
GAGACAGCTGCTCTACTGGTTCTT

<210> SEQ ID NO 535

<211> Length : 52

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 535
>HUMPHOSLIP_PEA_2_node_23
GAAGGTGTATGATTTTCTCTCCACGTTCATCACCTCAGGGATGCGCTTCCTC

<210> SEQ ID NO 536

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 536
>HUMPHOSLIP_PEA_2_node_24
CTCAACCAGCAG

<210> SEQ ID NO 537

<211> Length : 85

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 537
>HUMPHOSLIP_PEA_2_node_25
GTGTGGGCAGCGACAGGTCGCAGGGTGGCAAGGGTGGGCATGCTCTCACTTTGAGAAGGCCCTGACTCTGGCTCCCA
CCTCGCAG

<210> SEQ ID NO 538

<211> Length : 64

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 538
>HUMPHOSLIP_PEA_2_node_26
ATCTGCCCTGTCCTCTACCACGCAGGGACGGTCCTGCTCAACTCCCTCCTGGACACCGTGCCTG

<210> SEQ ID NO 539

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 539
>HUMPHOSLIP_PEA_2_node_29
TGCGCAG

<210> SEQ ID NO 540

<211> Length : 85

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 540
>HUMPHOSLIP_PEA_2_node_30
TTCTGTGGACGAGCTTGTTGGCATTGACTATTCCCTCATGAAGGATCCTGTGGCTTCCACCAGCAACCTGGACATGG
ACTTCCGG

<210> SEQ ID NO 541

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 541
>HUMPHOSLIP_PEA_2_node_33
GGGGCCTTCTTCCCCCTGACTGAGAGGAACTGGAGC

<210> SEQ ID NO 542

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 542
>HUMPHOSLIP_PEA_2_node_36
GTGCCCCACGACCTGGACATGCTGCTGAGGGCCACCTACTTTGGG

<210> SEQ ID NO 543

<211> Length : 15

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 543
>HUMPHOSLIP_PEA_2_node_37
AGCATTGTCCTGCTG

<210> SEQ ID NO 544

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 544
>HUMPHOSLIP_PEA_2_node_39
AGCCCAGCAGTGATTGACTCCCCATTGAAG

<210> SEQ ID NO 545

<211> Length : 87

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 545
>HUMPHOSLIP_PEA_2_node_40
CTGGAGCTGCGGGTCCTGGCCCCACCGCGCTGCACCATCAAGCCCTCTGGCACCACCATCTCTGTCACTGCTAGCGT
CACCATTGCC

<210> SEQ ID NO 546

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 546
>HUMPHOSLIP_PEA_2_node_41
CTGGTCCCACCAGACCAGCCTGAGGTCCAG

<210> SEQ ID NO 547

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 547
>HUMPHOSLIP_PEA_2_node_42
CTGTCCAGCATGACTATG

<210> SEQ ID NO 548

<211> Length : 27

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 548
>HUMPHOSLIP_PEA_2_node_44
GACGCCCGTCTCAGCGCCAAGATGGCT

<210> SEQ ID NO 549

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 549
>HUMPHOSLIP_PEA_2_node_45
CTCCGGGGGAAGGCCCTGCGCACGCAGCTGGACCTGCGCAG

<210> SEQ ID NO 550

<211> Length : 43

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 550

>HUMPHOSLIP_PEA_2_node_47
GTTCCGAATCTATTCCAACCATTCTGCACTGGAGTCGCTGGCT

<210> SEQ ID NO 551

<211> Length : 15

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 551
>HUMPHOSLIP_PEA_2_node_51
CTGATCCCATTACAG

<210> SEQ ID NO 552

<211> Length : 49

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 552
>HUMPHOSLIP_PEA_2_node_52
GCCCCTCTGAAGACCATGCTGCAGATTGGGGTGATGCCCATGCTCAATG

<210> SEQ ID NO 553

<211> Length : 83

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 553

>HUMPHOSLIP_PEA_2_node_53
GTAAGGCTGGGGTGTGAGGATGGAGGAAGAAAGGAGGGGTGAACTGGGCGGGCCCAGACTGAGCGGGGTGCTCCCAC
CCACAG

<210> SEQ ID NO 554

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 554
>HUMPHOSLIP_PEA_2_node_54
AGCGGACCTGGCGTGGGGTGCAGATCCCACTACCTGAGGGC

<210> SEQ ID NO 555

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 555
>HUMPHOSLIP_PEA_2_node_55
ATCAACTTTGTGCATGAGGTGGTGACGAACCATGCG

<210> SEQ ID NO 556

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 556
>HUMPHOSLIP_PEA_2_node_58
GGATTCCTCACCATCGGGGCTGAT

<210> SEQ ID NO 557
<211> Length : 36
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 557
>HUMPHOSLIP_PEA_2_node_59
CTCCACTTTGCCAAAGGGCTGCGAGAGGTGATTGAG

<210> SEQ ID NO 558
<211> Length : 23
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 558
>HUMPHOSLIP_PEA_2_node_60
AAGAACCGGCCTGCTGATGTCAG

<210> SEQ ID NO 559
<211> Length : 9
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 559
>HUMPHOSLIP_PEA_2_node_61
GGCGTCCAC

<210> SEQ ID NO 560

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 560
>HUMPHOSLIP_PEA_2_node_62
TGCCCCCACACCGTCCACAGCAG

<210> SEQ ID NO 561

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 561
>HUMPHOSLIP_PEA_2_node_63
CTGTCTGAGCCCTCAATCCCCAAG

<210> SEQ ID NO 562

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 562
>HUMPHOSLIP_PEA_2_node_64
CTGGCAG

<210> SEQ ID NO 563

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 563
>HUMPHOSLIP_PEA_2_node_65
CTGTCATTCAGGACCCCAAC

<210> SEQ ID NO 564

<211> Length : 90

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 564
>HUMPHOSLIP_PEA_2_node_66
CCCTCTCAGCCCCTCTTTTCCCACATTCATAGCCTGTAGTGCCCCCTCTAACCCCCAGTGCCACAGAGAAGACGGGA
TTTGAAGCTGTAC

<210> SEQ ID NO 565

<211> Length : 22

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 565
>HUMPHOSLIP_PEA_2_node_67
CCAATTTAATTCCATAATCAAT

<210> SEQ ID NO 566

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 566
>HUMPHOSLIP_PEA_2_node_69
CTGAGGAGCAAT

<210> SEQ ID NO 567

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 567
>HUMPHOSLIP_PEA_2_node_71
GCCCAGTCCCTAC

<210> SEQ ID NO 568

<211> Length : 105

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 568
>HUMPHOSLIP_PEA_2_node_72
CTTGATCGTGAGAAAGGCGATGTGGGAGAACTCCTTCACGAAGCCGGCAATCTGCTCCCCGCTGTCCCCGTACTTCA
CTAACCAGGGCCGGCGCTGCACCTCCAT

<210> SEQ ID NO 569
<211> Length : 108
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 569
>HUMPHOSLIP_PEA_2_node_73
CTGCCCCACCAGGAAAGACATCAGCCTACAGCAGCTGCATCCTTGCTCACAGCTACCAGCAAGACCTTAGGGCTGGG
AATTCCTCCACACTTGCCCTCTGTGGGCCAG

<210> SEQ ID NO 570
<211> Length : 90
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 570
>HUMPHOSLIP_PEA_2_node_74
AGCCAGGCAGCCAGCTGGCCACTCCCAGGCATACCCGCTCCCAATCCTCCACAGCAGCCCCTATCCCAGGGCCAGGA
ATCTCTACCTTAC

Segment nucleic acid sequences:

<210> SEQ ID NO 571

<211> Length : 774

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 571

>AI076020_node_0

CGCTCAGTCCGGCAGCGCAGCAGGAGGGAGGCGCGAGGCAGGAGCCGGCGGCTGGGCTCCGCAGCGCAGCCAGCGCA
GCGCGGCGCCCCGGGCCCGGCCCCATGCCCGCAGCCCCGCCGGACCGTCCTTGAGCGCGGGCGCCTAGCCCGCGCCC
CCTGCCCGCCGGCACCATTGCCCCGACGGCGCGGCCGGGCGGCCCGGCGCTCCCCAGGCTCCGCGCGGGCCGAAAGA
CGCTGCTAGCGGCCGCCGCGGGTGTGGTGATGCTGCTGGTGCTGGTGGTGCTCATCCCCGTGCTGGTGAGCTCGGGC
GGCCCGGAAGGCCACTATGAGATGCTGGGCACCTGCCGCATGGTGTGCGACCCCTACCCCGCGCGGGGCCCCGGCGC
CGGCGCGCGGACCGACGGCGGCGACGCCCTGAGCGAGCAGAGCGGCGCGCCCCCGCCTTCCACGCTGGTGCAGGGCC
CCCAGGGGAAGCCGGGCCGCACCGGCAAGCCCGGCCCTCCGGGGCCTCCCGGGGACCCAGGTCCTCCCGGCCCTGTG
GGGCCGCCGGGGGAGAAGGGTGAGCCAGGCAAGCCGGGCCCTCCGGGGCTGCCGGGCGCGGGGGGCAGCGGCGCCAT
CAGCACTGCCACCTACACCACGGTGCCGCGCGTGGCCTTCTACGCCGGCCTCAAGAACCCCCACGAGGGTTACGAGG
TACTCAAGTTTGACGACGTGGTCACCAACCTAGGCAACAACTACGACGCGGCCAGCGGCAAGTTTACGTGCAACATT
CCCG

<210> SEQ ID NO 572

<211> Length : 170

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 572

>AI076020_node_3

GTGCGGGCCAGTGCTATTGCCCAGGACGCGGACCAGAACTACGACTACGCCAGCAACAGCGTGATCCTGCACCTGGA
CGCCGGCGACGAGGTCTTCATCAAGCTGGATGGAGGCAAAGCACACGGCGGCAACAGCAACAAATACAGCACGTTCT
CTGGCTTCATCATCTA

<210> SEQ ID NO 573

<211> Length : 175

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 573
>AI076020_node_8
CGGCGGCGGGTTTGCTTCCTGCGCTCTGAGATGAGCTGCCCTCGGCTCCCTCCGGGGGTGGCGCGCCCGGGGGAGGGG
GGAGTTGGGGGCTGGATAGCTTCCCAGCACCCTCAGAGCCCCCGCCCGGCTGTGCCCCGTCTGACCAAAGTTATAAT
AAAAACATTTTCACCCCGCAG

<210> SEQ ID NO 574

<211> Length : 83

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 574
>AI076020_node_1
GCACCTACTTTTTCACCTACCATGTCCTCATGCGCGGCGGCGACGGCACCAGTATGTGGGCAGACCTCTGCAAGAAT
GGCCAG

<210> SEQ ID NO 575

<211> Length : 102

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 575
>AI076020_node_4
CTCCGACTGAGCTCCCCACGTCTCCCTCCACCCACGTCCCTCACCCGCCGGGGTCCCCTCCGGGCGGGGCAGACGAT
GACTCGCCCCTCGCCCACCCGCTCG

<210> SEQ ID NO 576
<211> Length : 115
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 576
>AI076020_node_5
CTGCCCGGCCCTCCCCGGCTATGACGCCCCCGGCCCGTGCTCAACACCGCCTGGGCCACAGCTAGGCCCTCCCACCG
GCTCGCTGCAGAGCCGGGCCCAGCGCGCCCTGTCCCCG

<210> SEQ ID NO 577
<211> Length : 76
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 577
>AI076020_node_6
TGCCAGGGAACCGGGGTTGACCGCCCCCGCCCAGCCCGCGCTATATATTTGTACAATAGGACTGTTTACTGCCCAC

<210> SEQ ID NO 578
<211> Length : 38
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 578
>AI076020_node_7
CTCCGCCTGCCAGCCCACCCCAGCCTGGGGAGAGGTCG

Segment nucleic acid sequences:

<210> SEQ ID NO 579

<211> Length : 1,098

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 579
>T23580_node_17
TTTTTGTTTGTGTACTTAATAAAGGGTAAATATGTCATGTTTGTTTGGAACAGTCATGGTTAATATCTATGTTGTCC
CAGTATATCTATTAATAGAACTCTCTTTCACTCTCAACAGCGTCCTAGTCCGGATGACAAATTATATGGTTATCTCT
CAGTAAAGGGTCTTTTTTTAAAATGATTTTTTTTTCAGGGGGTAGGGTAGGCAGGAAGCTTAAACTGGGTAATTTAG
TTGTAGAAGAGTGCCCTGTGGCAAATAATTGATTATTCATTTCCAGCATCCCCTTTTCTTCTCCTTGACAGTTATTA
AAAAAAAAAAAGTTACCAGCTTATGTCATTTTAAAGAACACTCGCCCTGAAAACTTCTGAGAGGTTGGCCATTTGAA
ACCCTGGTTTTAGTGTCTGTATTATTAGTGAACTACCGTGTTCCCATGTGGCTACACAACCACAATTATGTACTATC
TGGCTCTTTACCAAAGTTTGCAGACCTCTAATCTAGAGTGCGACATTTCCCCTCATTAACTCTTAGGTCCCTTGGCT
CTAAAAGGGTATATTCATCTTGGCCCTATACAGGGAAAGGGGGAATGGGATTAATGATGTGCTTTGTAAGAAGAACC
AATTTTAATTTTCACAAAGGCTTGACGTAGCTGTGAGAGAAAGGGTAAGAAGAAGCAGGCTTCTTCTTAGAAGTCTG
AGATGGCCTAAAGTGGTGGGGGAAGAAGGGAGAGTGGGGAGAAAGAGAAACAAGAAAAGCTGAGAGTGAATTCCCCA
GAGAGGTAGCCACTGATTCTGCCCTACTCTTTGCTGGAATTCTGGAAAACACCTGGGCTTCTAAAAGATAGGGAGCT
CATGCATCATGGTAGGGCCACAGCTCAGGCTAGGGCCAGAGATAGCTCAGAGTAGCGCCACGGCTCAGGGTAGGTCC
ACAGTGCAGGGTAGGGCCATAGCTCAGAGTAGGGCCATAGCTCATAACCACAGCTCAGGGTAGGACCTGCTGATCTA
TTTGGGGACCCCCAGCAGAGCCTGTCTAATTGCATATCTTGAAAAGGATTGGAAAACTGTCATAATGACATTATTCC
CTCTCACTTTCCTTGTCCAG

<210> SEQ ID NO 580

<211> Length : 259

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 580
>T23580_node_18
GAAAGCCCAGCAGAATCGGAGAGGCTTTTCCGAGGAGCAGCTTCGCCAGGGACAGAACGTAATAGGCCTGCAGATGG
GCAGCAACAAGGGAGCCTCCCAGGCGGGCATGACAGGGTACGGGATGCCCAGGCAGATCATGTAGGACGCGGCATCC
TGCCCCTGGTAGAGAGGACGAATGTTCCACACCATGGTCTCTACGAAAAAGAAATAGTTAGTCACCTTCTGACCTTC
TCCTCTTTCTCAAAGCCTTCTGTCCCTG

<210> SEQ ID NO 581

<211> Length : 201

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 581
>T23580_node_21
CCGAGAATCCGCGTTGCCTACTGCTGCCACCTCCTGTTCATTTAGAACTATGCAAAGACTCCGCTTCCGTTTTCCTG
AGCTCCTCGGGCCCCAGAGTCTCTGTTTGATTATTTATTTATTTATTTATTTATTTGCCAAAAATTCTCCTCTTCAA
CTTATAGAATGCACCTAATAAAGTAATTAGTCTTGTGTCTTACAGTG

<210> SEQ ID NO 582

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 582
>T23580_node_19

GTTTTTGCAAGTG

<210> SEQ ID NO 583

<211> Length : 11

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 583
>T23580_node_20
CTGCATTTCCG

<210> SEQ ID NO 584

<211> Length : 128

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 584
>M79217_PEA_1_node_2
ATTTTCAAACACCGCATTCCCCTTAAACCTCCCAGGCTCCCATCACTCCCAGCAAAGAGCCAGCCCGCTACTTTATG
GAGACAGGAGAGACTGTCAGACAGTTCCTGCTGCTGCCCAGGTGCATCTGA

<210> SEQ ID NO 585

<211> Length : 177

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 585

>M79217_PEA_1_node_4

GGCGCCCTTCGGCAAGTTCCGCAGTCGCCTGTCGGAAATGGCTGCCGGCCGGCAGGGGGAGCGGCGGATCAGGCGCG

GCCTGGAAGGCGGGCGGCCGGCAGCCAGAACGGCTTCTGGGACGCCGACTTTCGCGCAGGCGGCGGCGGCGGCGGCG

GCGGGTCCCTGAGCTGGAAGCCG

<210> SEQ ID NO 586

<211> Length : 597

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 586

>M79217_PEA_1_node_9

TGATGTCTGCTAATGGAAGATAAATGAGAAGCAAACTGGGAAATACATTTTGTCTCAGGATTACTGCATCTACTACT

GGATAAAGATCAAAAGAGTTCTCTTCAACCCTTTCAACTCTACATTTAACAAATTGAGCTTTTCAGAGTCTTTTTTT

GTAAAGTATTTCCAAAGAAGACTTATAGGTTAGGAATAAACATAAACTACCCAGGTTGGCTAGGAAGGTATTTCTGT

TCATCTAAAGATGATGCCCAGGTGTGGAACAGGATAAGAAAAGACCATGGACATCTTTGTCCCATGAATTTAGTTGG

TCATCGTGTTACAGGGCTATAATGCCGCTCTAGATCCAGTTAAATAAGAAGTGGGGAGGGGTTGTAAACTGCAGCTT

TTTGGGGCACTTATCCATTTATTACCCCAAGTAAAAGACCTATACCAAACAGCAAACAACATCTCTGCATTGTCATT

ATAATGTTCTTTGAGACACAGCCAGTGTTCCAGCCATTGTTCCATCTAAGATTTAAGCATTTTCTAGAAATGTATGG

TGGCAGGGGTGTTGAACATAACTTCTTCAAGACTGACATGGTTCTCTTTCTTTTGCAG

<210> SEQ ID NO 587

<211> Length : 483

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 587

>M79217_PEA_1_node_10

GCCTGATTGTTGGCAAAGGCATCATAAGAAGCTGGCATTTATTTCTGTTCTAACCTATTACTGTATAACTGTGAATA
GACACTATGCATATTTGTTGGTCAGCAAAACCAAGAAACAAGAGCTATGGCATTTGAAAAAGTCTGTCTGATTCCAG
GGTGTTTTTCCTGGGTTTCATCATCAGGTACCTCCTCCCTTTCATCTCAGCAAGAATGTGGCACCTTTTATCGTTTG
ATAAAGATTAAGGACATGTTCTTTGGTCAACAGCCAGAACTTAAAATCTGCTGGAATAGGGTCAGAGACCATTTCAG
CTGCAGCTGAGGAAAATGAAATGTTCATTTTATTTGGTGCCTTGTCTGGGGAGCACACTAACTCTTCTGGAAACGTG
TCAGTGAAACAGAGATCGTTTTGTGGAATAGCTAACCCATGGTTATGGCGAGTGACCCGACGTGATCTGGGGGGCAG
GCTGCAGAGGACTCATGACAG

<210> SEQ ID NO 588

<211> Length : 443

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 588

>M79217_PEA_1_node_11

GCTATACCATGCTGCGGAATGGGGGCGCGGGGAACGGAGGTCAGACCTGCATGCTGCGCTGGTCCAACCGCATCCGC
CTCACGTGGCTCAGCTTCACGCTCTTTGTCATCCTGGTCTTCTTCCCGCTCATCGCCCACTATTACCTCACCACTCT
GGATGAGGCTGATGAGGCAGGCAAGCGGATTTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAAGCACGTGC
TGGATCTGTGCCGCATCCGGGAGTCGGTGAGTGAAGAGCTCCTGCAGCTGGAGGCCAAGCGCCAAGAGCTGAACAGC
GAGATCGCCAAGCTGAATCTGAAGATCGAAGCCTGTAAGAAGAGCATTGAGAACGCCAAGCAGGACCTGCTCCAGCT
CAAGAATGTCATCAGCCAGACCGAGCATTCCTACAAGGAGCTCATGGCCCAGAACCAG

<210> SEQ ID NO 589

<211> Length : 528

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 589

>M79217_PEA_1_node_13

CTGCTCCCAGAGAAGGACGATGCCGGCCTCCCTCCCCCGAAGGCCACTCGGGGCTGCCGGCTACACAACTGCTTTGA
TTATTCTCGTTGCCCTCTCACCTCTGGCTTCCCGGTCTACGTCTATGACAGTGACCAGTTTGTCTTTGGCAGCTACC
TGGATCCCTTGGTCAAGCAGGCTTTTCAGGCGACAGCACGAGCTAACGTTTATGTTACAGAAAATGCAGACATCGCC

TGCCTTTACGTGATACTAGTGGGAGAGATGCAGGAGCCGGTGGTGCTGCGGCCTGCTGAGCTGGAGAAGCAGTTGTA
TTCCCTGCCACACTGGCGGACGGATGGACACAACCATGTCATCATCAATCTGTCACGTAAGTCAGATACACAGAACC
TTCTCTATAACGTCAGTACTGGCCGTGCCATGGTGGCCCAGTCCACCTTCTACACTGTCCAGTACAGACCTGGCTTT
GACTTGGTCGTATCACCGCTGGTCCATGCCATGTCTGAGCCCAACTTCATGGAAATCCCACCACAG

<210> SEQ ID NO 590

<211> Length : 1,146

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 590

>M79217_PEA_1_node_14

GTGCCGGTGAAGCGGAAATATCTCTTCACCTTCCAGGGCGAGAAGATTGAGTCTCTGAGGTCTAGCCTTCAGGAGGC
CCGCTCCTTCGAAGAGGAAATGGAGGGCGACCCTCCCGCCGACTACGATGACCGGATCATTGCCACCCTGAAGGCGG
TGCAGGACAGCAAGCTGGATCAGGTCCTGGTGGAATTCACCTGCAAAAACCAGCCCAAACCCAGCCTGCCAACTGAG
TGGGCACTGTGTGGAGAGCGGGAGGACCGCTTGGAATTGCTGAAGCTCTCCACCTTCGCCCTCATCATTACCCCCGG
GGACCCTCGCTTGGTTATTTCCTCTGGGTGTGCAACACGGCTCTTCGAAGCCCTGGAAGTCGGTGCCGTCCCGGTGG
TGCTGGGGGAGCAGGTCCAGCTTCCCTACCAGGACATGCTGCAGTGGAACGAGGCGGCCCTGGTGGTGCCAAAGCCT
CGTGTTACCGAGGTTCATTTCCTGCTCAGAAGCCTCTCCGATAGTGACCTCCTGGCTATGAGGCGGCAAGGCCGCTT
TCTCTGGGAGACTTACTTCTCCACTGCTGACAGTATTTTTAATACCGTGCTGGCTATGATTAGGACTCGCATCCAGA
TCCCAGCCGCTCCCATCCGGGAAGAGGCGGCAGCTGAGATCCCCCACCGTTCAGGCAAGGCGGCTGGAACTGACCCC
AACATGGCTGACAACGGGGACCTGGACCTGGGGCCAGTGGAGACGGAGCCGCCCTACGCCTCACCCAGATACCTCCG
CAATTTCACTCTGACTGTCACTGACTTTTACCGCAGCTGGAACTGTGCTCCAGGGCCTTTCCATCTTTTCCCCCACA
CTCCCTTTGACCCTGTGTTGCCCTCAGAGGCCAAATTCTTGGGCTCAGGGACTGGCTTTCGGCCTATTGGTGGTGGA
GCTGGGGGTTCTGGCAAGGAATTTCAGGCAGCGCTTGGAGGCAATGTTCCCCGAGAGCAGTTCACGGTGGTGATGTT
GACTTATGAGCGGGAGGAAGTGCTTATGAACTCTTTAGAGAGGCTGAATGGCCTCCCTTACCTGAACAAGGTCGTGG
TGGTGTGGAATTCTCCCAAGCTGCCATCAGAGGACCTTCTGTGGCCTGACATTGGCGTCCCCATCATG

<210> SEQ ID NO 591

<211> Length : 128

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 591
>M79217_PEA_1_node_16
GTGGTCCGTACTGAGAAGAACAGTTTGAACAACCGATTCTTACCCTGGAATGAAATTGAGACAGAGGCCATCCTGTC
CATTGATGACGATGCTCACCTCCGCCATGACGAAATCATGTTTGGGTTCCG

<210> SEQ ID NO 592

<211> Length : 145

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 592
>M79217_PEA_1_node_23
GGTGTGGAGAGAAGCTCGGGACCGCATCGTGGGCTTCCCTGGCCGTTACCACGCATGGGACATCCCCCATCAGTCCT
GGCTCTACAACTCCAACTACTCCTGTGAGCTGTCCATGGTGCTGACAGGTGCTGCCTTCTTTCACAAG

<210> SEQ ID NO 593

<211> Length : 412

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 593
>M79217_PEA_1_node_24
GTAAGAAAAAGCTGGTAATAATGGCATCGACTTGGTGAGAGTTTCACCTTTGTGTGGTAGCGGAATGCTGCCCTCAG
CTTAGCTCTCCTAACGCTTCTTACATGTTTCTTTTGTGCTAGAAGTCAGTTTTTTCTATTTTTACAGACAATGATCA
AGATGCTTAGAGCAACTCTGGGATAAAAAGTCAAGATGAGAGGGCTGCCTGTACAGTTGCACATAGGCCATTTGGAA
ACCACTTTATCTTTCTGGGCGTTGGCTCTCCGTTTGTAAAACTGAGGGCACTGGGCTAAAGACACCTCAAATACCTT

CCAGTTTTAACACTGCCACCCTAGATATGGCCCAGCCATCAGAAGGTGACCTGGGCACTTTTCTGACTTAGATATAC
CATGCCTGTCCCGGGCCCCACGATGAG

<210> SEQ ID NO 594

<211> Length : 245

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 594
>M79217_PEA_1_node_31
CTTCTTCGTGAAGGTGTACGGCTACATGCCCCTCCTGTACACGCAGTTCAGGGTGGATTCTGTGCTCTTCAAGACAC
GCCTGCCCCATGACAAGACCAAGTGCTTCAAGTTCATCTAGGGGCAGCGCACGGTCTGGGGAAGAGGATGAGCAGAG
GGAGGAAGATGGCTCCCAAGGTTCCTAGGCATTGCAGGACCTTGGGCACATCTGCTGGTGGGTGGCCCAGAGCCTCT
GCTGGAAGGGGCAG

<210> SEQ ID NO 595

<211> Length : 617

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 595
>M79217_PEA_1_node_33
CTGGAGCCCTGGGCGGAGTCCCCGGGGTTCCCCACACAGGGCACTGACTGATAGCTTACACTGAGGACTGTGGCGAC
TCTGCAGAGTCACTCACACCGTTCGTACGCCCAGGACAGCTGGTTCGTGGTTTTTACATTCAATAACAACTATTATG
ATTATTTAAAAAGAGAAAGTTTCAGATTTGCCATTCAAGGCTTATTTATATATATGTGTGTGTATATAAATACATGC
ACACACTTGCATACATATATATTTTTGGCTGGGGGAGTGTGAGTTTTGCCTTTCTAAGGGAGGGACCGCGCAGGCTC
CTTTGTTCTGTATTCTGGCGGAGATGGGTCCTGGCCTTGTGTCACTGGCTTATCCTTAAAGATCATCTCCCATCCTC
CCCAGCGCCATCTGTGTGCAGCAACCAGAAAGGGATGAACTTGGCCCTCTTGCGGGCCTGGACAAGGTCTCTTCCTT
ACCCTTTCTGTTGCCAGTCAGCAACCTGTAACTCACATTCTCTTCCCAGTGAATCCCTGGGAGCGCCTGACCCTGGT
GGGCTGTTCAGCTTCCTGCTGCTGGGGCCAGCGATTTTTGAGGATTTATCTTTAGGCCAGGCTTGCCTCCGTACTTA
T

<210> SEQ ID NO 596

<211> Length : 238

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 596
>M79217_PEA_1_node_34
CCCTGCTCTCCCATTTCTCTCTTGTTTGAGAGAGAATGAGGAAGCAAAGAGTGAGAAAGAATAGGGGCTGAAGACGC
CACTCCCAGATGGCTCTTTCTATCCTGCTCTTCTGTTGAAACACACGTGCTGTGGGCCTCAGGCGTTTCTGAAGTGC
TCTTTCTTGGATTGGACAGGAGATCAGCAGCGTGCACATCTGCTGTGGTCTGAAGTGGTTTGCAGGTCAGCCTCCTC
TCCCTAG

<210> SEQ ID NO 597

<211> Length : 128

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 597
>M79217_PEA_1_node_35
TGTAGAGCAAGCCAGTGTCCTTCGAGGAACCCACCCGGCTGGCCGGGAAGTTTTACAGCAAGGCGCCTGCCTTGGGA
TAATTCCTTGGTGAAATTCACCTTCCCCCCGCCTCTGTCTGGAGCCCCATC

<210> SEQ ID NO 598

<211> Length : 242

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 598
>M79217_PEA_1_node_37
CTGTAGGACTCCCCGAGGTTTGGTATGTGCTAGAACAATGGGAGGCTGTGATTTGCTGTGTAAGCTCACATCCAGCC
TTGGAATCTAACGGGCATTCACAACCCGAGTTACCACTTTCCACTCCCTGCTTAGGATTCTGTTCCCTGGGCTGAAA
CTGAAATAAGCTAATTTTTTGGGTCACGGTGGCAGTAGGGGAACCTAGGAGGGTGTGAGTGGCATTTGTCAGGGATT
TAGCCCATGAC

<210> SEQ ID NO 599
<211> Length : 156
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 599
>M79217_PEA_1_node_38
GTGTTTCTTGAACCCTACTTTCTGGAAGTGGAGTTGACTCTGGAAGTTTTCTAGCAACTGAACAAAAGCTCAGGTTT
GTCCTGGTCATGCACATGCCTTAAGCCAGTTCCGTCTTCCCTAGACCTTGGCATCCTGTGCTTCTATTTCTTGGAAT
AC

<210> SEQ ID NO 600
<211> Length : 730
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 600
>M79217_PEA_1_node_41
CCTGTAGCTCCCCTGCTGAGGCAGTGTGGTTATGTTCCCAGCAGTGGGGGTCAGACGCCCTTCCTCAGAACTTTCTA
GTTGCCCTCTACCTGACTCCTGACTTGTATTCCTTTTAGCAGTAGCCTTCTTCCCTCGGGGAGCCAAAGAGTGTGGT
GTGTGGCGCTATATTGTGGCTGCTATTTCATCTGGTTTCTTTTAATGTGAGGAACTCACATACTGACTTCAGTGGGA
CTCGGTGAGCCGGGGCCGTCTGTGTGGTGGGACCCCCTTTAGCGGGACTCAGTGAGCTGGGGCCGTCTGTGTGGTGG
AGCCAGGGCCTCTCCCTTTAGTGGAGCCAGGTTGTCGGGCCCCGAATGTCACTGGTGGATCTAAGAAGGGCTGAGTG
GTCTGACACCAAAACATGCCGCAGGGAGGGCTGTGGTGCCGGTGCTTCCAACAAGGACAGCCCTCCTTGACCCTGAA
AGGAACACTGGCTTGAAGGACTGCAGACAGGCTCTGAGGGGCACGCCCTCCTCAGCGAGAGGCAGCAAGGTGGCCAC

AGTGTCACTGGTCAGGTGCTTCTCACCACGGGAAAGCCGCCGACCTGTGACTCGCTTGAGATGGGAAAGCGGCGCCA
CAGACCCCGGGTCTCCTTGGCTGTCTGTGGGCCGCCCCTGGCCACCTTGTCCTGGCTCGCAGGGTGCAGGAGCGCCT
CGTTCTCTGGGTGGCCGGCTTGCTGCTCCGGTTTGGG

<210> SEQ ID NO 601

<211> Length : 188

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 601
>M79217_PEA_1_node_44
CTGTCCGGAAGGATGGGATCTTTCTGGGAGCTGCGCCGGACAGAGTGGGGAGCTCCTAGTTTGTGGGGGGAAGCTTT
GATATCCATGCCACGTCCATCCACCCCACCCCTTTTCGTCACGAGCACAATGGTCTTACATTGGATTTTTGTAAAAA
AATAAAAATAAATGGAGACTTTAACTCAAGCAGC

<210> SEQ ID NO 602

<211> Length : 49

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 602
>M79217_PEA_1_node_0
GGCTGCAGTGGCCGCCGCTGGAGACCGCGGGACCTGGCGCTGCAGCGAG

<210> SEQ ID NO 603

<211> Length : 94

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 603
>M79217_PEA_1_node_7
GAGAGCAAGCCCTGGAGGTTCACTCTTTCAAGAAGTCGTGTGCTGAGGTGTAATGCTACACAAGTCAGAGGAAGGAA
GGGTCCTGAAACACATG

<210> SEQ ID NO 604

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 604
>M79217_PEA_1_node_12
CCCAAGCTGTCCCTGCCCATCCGA

<210> SEQ ID NO 605

<211> Length : 79

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 605
>M79217_PEA_1_node_19
TCCTGAGCTCAGGCAACCTGCCCGCCTCGGCCTCCCAGAGTGCTGGGATTACAGGCATGAGCCACGGTGCCCAGCCC
AG

<210> SEQ ID NO 606

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 606
>M79217_PEA_1_node_21
ATGGGTTCTCACTTTATTGTTCAAGCTGGTCTGAAACTCCTGGCCTCGAGCAAGCCTCCCAAGTGCTGGGATTACAG

<210> SEQ ID NO 607

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 607
>M79217_PEA_1_node_26
TATTATGCCTACCTGTATTCTTATGTGATGCCCCAG

<210> SEQ ID NO 608

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 608
>M79217_PEA_1_node_27
GCCATCCGGGACATGGTGGATGAATACATCAACTGTGAGGACATTGCCATGAACTTCCTTGTCTCCCACATCACTCG
GAAGCCCCCCATCAAG

<210> SEQ ID NO 609

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 609
>M79217_PEA_1_node_30
GTGACCTCACGGTGGACATTCCGATGCCCAGGATGCCCTCAGGCCCTGTCTCATGATGACTCCCACTTCCACGAGCG
GCACAAGTGCATCAA

<210> SEQ ID NO 610

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 610
>M79217_PEA_1_node_32
CAGGAGGAGTGGAAGGAAACCGCTGCCTTTATCTTGAAGTCAGCCACACTGGGC

<210> SEQ ID NO 611

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 611
>M79217_PEA_1_node_36
CTGTGTTATCTGTGGTTTTTGGACCCCTAATGTCAGCTTGG

<210> SEQ ID NO 612

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 612
>M79217_PEA_1_node_39
GTTCTCCTCTGACCTGCCTGTACCACGTGGGTCCTCTTCAAGTACTGTTTTGAAGCTGGGCTCTTTTGTGTAGCTCC
CACCCAC

<210> SEQ ID NO 613

<211> Length : 107

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 613
>M79217_PEA_1_node_40
CTGTAGGGCTAGCTCGGCTTAAGGGAACTCTCCCCATTGGCAAACCGGACCCGGCCGCCGCCAGGACTGTGTTTCCA
AAGGTTCCCCGCCCCCAACCCCAGCATCAG

<210> SEQ ID NO 614

<211> Length : 86

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 614
>M79217_PEA_1_node_42
CTGTCTTACCATAACACCGTCCCAGGGCTCTGCAGGCCACTGTGAGCGCTGGCTCCCTGGGCAGTGCTCCTCCGTGT
GGACTGTGC

<210> SEQ ID NO 615

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 615
>M79217_PEA_1_node_43
CTCAGGCCAGGGCTCACCAGCTGGGGTC

Segment nucleic acid sequences:

<210> SEQ ID NO 616

<211> Length : 355

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 616
>M62096_PEA_1_node_0
CGGCAGAGCGCGCTGGTGCTGATGCAGGATGGCTGAGCGCGCAGGAGCCCGGGAGGTCTGAGCCGGGCGAGGCTCGC
TCCCTGCGCATCGCCTCCTCCGCCCGCCGCGTGGTCGCGGGCAGGTGGGCCGGGGGGGCGCTGGGCAGGGGCGGGGCA
GGGCCAGGGCAGGCCGGTCTGCAGCCGGAGGGGCCGGAGCGGAGAAGCTGCCCACCTTCCCGGGCTCGGAGCGGCCG
GGGCTGCTCAGCCGGCCGGGCTCGCGATGACCTGCTGAGAAGCGTCGTCGGAGGCTGCAGGAGGCGGCCTAGCTGTG
GGCGGTGCAGCTCGCGGCCTCCTCCCTCGTCGTTCCCGGCCCCGGCC

<210> SEQ ID NO 617

<211> Length : 148

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 617
>M62096_PEA_1_node_2
CCCCTCCCTACCGCCGGCCGAGATGGCGGATCCAGCCGAATGCAGCATCAAAGTGATGTGCCGGTTCCGGCCCCTCA
ACGAAGCGGAGATCCTCCGCGGGGACAAATTCATCCCCAAATTTAAAGGCGATGAGACCGTGGTGATCGGG

<210> SEQ ID NO 618
<211> Length : 125
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 618
>M62096_PEA_1_node_15
ACATGAATGAACACAGCTCTAGAAGTCACAGTATCTTCCTGATAAATATTAAACAAGAGAATGTAGAGACTGAAAAA
AAACTCAGTGGGAAACTTTATTTGGTTGATTTGGCTGGGAGCGAAAAG

<210> SEQ ID NO 619
<211> Length : 147
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 619
>M62096_PEA_1_node_17
GGTTTGACTTAGCCTGTGGCAGAGGCAGCTGCACACATGTGGGAGGGATATGAACTTCTGAGAAAAGAGAAAATCCC
ATGTTGTACCCGACTCTAATAGGCCAGGAACGAGCTTTGCCATTGGAATGTGGCAGTCCTGCCTCTGCAG

<210> SEQ ID NO 620
<211> Length : 125
<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 620
>M62096_PEA_1_node_19
GCTGATTGTCCCCATGAAGGCCAGCCTTGAAGCTTGGTCAGTCTCCCTAACTGTATGATTGATCCCCACTTATTGCA
CTACATCACTGAGTTCCCGTATGCCAAGTTATGGCCACTTACATCCAC

<210> SEQ ID NO 621
<211> Length : 149
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 621
>M62096_PEA_1_node_23
AAAACACATGTGCCATACCGGGACAGCAAGATGACTCGGATTCTTCAGGACTCTTTGGGTGGGAACTGCAGAACCAC
CATCGTCATTTGCTGTTCTCCTTCTGTCTTCAATGAGGCTGAGACCAAGTCCACACTGATGTTCGGACAGAG

<210> SEQ ID NO 622
<211> Length : 149
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 623
>M62096_PEA_1_node_27
AGCTAAGACCATCAAGAATACAGTCTCTGTGAACCTAGAACTGACAGCAGAAGAATGGAAGAAGAAATATGAAAAAG
AGAAAGAGAAAAACAAGACTTTGAAGAATGTTATCCAGCATCTGGAGATGGAGCTAAACAGGTGGAGGAATG

<210> SEQ ID NO 624

<211> Length : 167

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 624
>M62096_PEA_1_node_29
ATTTTCTAGCAGCACACGTGTTTGGAAAGCTACTAGAATAATTGAATAATTCAGCACCTGAGGCTGGTGGATGATTC
TTTGCAATTTGGCAGGAATGGGAGAGTCGGGAGCAGTAGTTGGCAAGGTGGGGAGTAGCCATATGAAGTTTTATTTC
GGGAATCCTCCAG

<210> SEQ ID NO 625

<211> Length : 176

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 625
>M62096_PEA_1_node_31
GAGAAGCTGTGCCTGAGGATGAACAGATCAGTGCCAAGGACCAGAAGAACCTGGAGCCTTGTGATAACACCCCCATC
ATAGACAATATTGCTCCTGTTGTTGCTGGCATCTCTACAGAGGAGAAAGAGAAGTACGATGAGGAGATCTCCAGTCT
CTACAGACAACTGGATGACAAG

<210> SEQ ID NO 626

<211> Length : 504

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 626
>M62096_PEA_1_node_34

GTAAAGAATGCAATATATTTTTTTTTCCACAAAGTTCTTCTATTACTCTTTGTTGTTGATGTTTGTTCCAGGAATTT
AATTGGCATAGAAGCTTTTCATAATTACAGAATCATGTGGAAATTTCTTGGTAGATGTCCCTTCACTGCCTCTTACA
AGCTGATTATCACTGAATTTAGAAAATAAATGTCTGACTTTCAAAAACCCCTGATGTTTTGAGATTGAGTAGCCAGT
GGCTACAGTTCGTTCTGGAAGGGCAGAGACCTTTGGTTGGGTGATCAAGCAAGGATGATCCTTTTTTATTTTTATTT
TTTTGAGACAGGGTCTCTCTGTTGTCCAGGCTGGAATGCAGTGGTGCAATCATGGCTCACTGCAACCTCCAGAGCTC
AAATGATCTTCCCGCCTAAGACTCTCAAGTAGCTAAGACTACAAGAATGTGCCACCATACCTAGCTAATTTTTTAAT
ATTTTGAGACAGAGTTTCTCTATGTTGGTCAGGGTGATCTTG

<210> SEQ ID NO 627

<211> Length : 207

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 627
>M62096_PEA_1_node_36
CTTTTAGCTTCCACAAGAAGAGACTATGAGAAGATACAGGAGGAGCTGACACGTCTCCAGATTGAAAATGAGGCAGC
CAAGGATGAGGTGAAAGAAGTTCTCCAGGCCCTGGAGGAGCTGGCTGTCAATTATGACCAGAAATCACAGGAAGTGG
AGGATAAGACCCGGGCCAATGAGCAGCTGACAGACGAGCTGGCCCAGAAAACG

<210> SEQ ID NO 628

<211> Length : 147

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 628
>M62096_PEA_1_node_38
ACTACATTGACAACCACACAGAGAGAGCTGAGCCAGCTACAAGAGCTTAGCAACCACCAGAAGAAAAGGGCAACTGA
GATCCTGAATTTGCTGTTGAAAGATCTGGGGGGAGATAGGTGGAATTATTGGCACCAATGATGTGAAAACT

<210> SEQ ID NO 629

<211> Length : 189

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 629

>M62096_PEA_1_node_40

TTGGCAGATGTGAATGGAGTCATTGAGGAGGAGTTTACCATGGCCCGCCTGTACATCAGCAAGATGAAGTCAGAGGT
CAAGTCCCTGGTGAACCGCAGCAAACAGCTCGAGAGCGCCCAGATGGACTCCAACAGGAAGATGAATGCCAGCGAGC
GGGAGCTGGCAGCCTGCCAGCTGCTCATCTCCCAG

<210> SEQ ID NO 630

<211> Length : 340

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 630

>M62096_PEA_1_node_48

GTGAGTCGGCCCAGGGCACCAGGGGTGTGGGGGTGGTCATGCCTCGGTCCTCTTGGGGAAGCCTGGAAGGATGTGGC
TCTTAGTCGAGGGCCCTGCTCACCTTGCCCTGTGGGCACTGCCCTGGTGACACAGCAGGCTGGGCGGGCTGCTTCCA
AGGTCTGTTCTCCGATCTGGAGCTGAGCCTCCTGGAGCCCTGGCATGGCAGGTGGCAGGCGGGCCCAGCTGCCTCTC
CTAGTCCCCGAGGGCCAGGGTCACATAGGTGATTCCGCTGGATGGACGCATGCCCCAGTAGATGGGGGGGAGCATCT
GTAATGAAAGCACCAAAAAAAAAAAAAAAAAA

<210> SEQ ID NO 631

<211> Length : 688

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 631

>M62096_PEA_1_node_50

AAGCTTCCCTGGCTTGTGCCTTTCAAATGGACTCTGGGTTTTCCTTCTGGTAGTGCATAGCTGTTCCTTTACAGGCG
CTTAGGCGTGGCTCTAGGAAAGGTTTTATGAGTCCTGGGCTGATGTAAATGTTGACCAAACACCCTCAACCAGATGG
CGAGTTTCTGTTTGCAGCAGAGCCCAGGCTGTCTTTTCTTCATAATTCTCTCTGTGCCCACTCCTCGAGGGCAGGAA
CTGTCCCTGTATCAGTGAGGCATTCGGACTTGGGAGATGTTTTTAGAACATCAGACCAGAAATGAGGGAAGGTGGAA
ATGGCCAAATCAGGTTCCCCAAGTGACTGCATGCCATCCGAGGGGCCGAGGAAGCAGAGTTCTTCTGACATGGGCTC
TCTGTTTTAAAATATCAGCCCTTCTCCCATCTCATTATTTTTCCCCTGAAGCTCTTGCACAAGCAAACTATAAATAC
ATCCTCAAAGCCTTATGTTTCATGACTCTTAGATGCACCCCAGAAATTAGTTTTCACTTGGGCACGGAGGAAGCTGT
GAGGCTGTTCTGTGATCCCTCCAAATCCTGCAGAATTACTGCCTTTATTGTACAGAGCTAATAGGGTTGGAACAGAA
CCACGGTTTTAGCCTGATGACTCAGAATTTCAGACTGATGTGGAATATATTGCTTTTTCCTCTCAATTTCAG

<210> SEQ ID NO 632

<211> Length : 136

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 632

>M62096_PEA_1_node_56

GTGAGTTCTCTTTGTCTGAATGGGACTGAGAAGAAAATCAAAGATGGCAGGGAAGAATCATTTTCAGTTGAAATATC
ACTTGCTTAAGTCGGGGCTGGTTATGCTTAAAAATTAATTACTGCACACCAGAGAATGT

<210> SEQ ID NO 633

<211> Length : 217

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 633

>M62096_PEA_1_node_60

CTGGTCCGGGACAAGGCAGACCTGCGCTGTGAACTGCCCAAGCTGGAGAAGCGGCTGCGTGCCACGGCGGAGCGCGT
CAAGGCTCTGGAGAGCGCGCTGAAGGAGGCCAAGGAGAACGCCATGCGGGACCGTAAGCGCTACCAGCAGGAGGTGG
ATCGTATCAAGGAGGCCGTGCGGGCCAAGAACATGGCCAGAAGGGCCCATTCAGCCCAGATCG

<210> SEQ ID NO 634

<211> Length : 1,320

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 634

>M62096_PEA_1_node_65

ATATGACTCCACGTAGCATGTCAAGGACTACATTAATCACCAATTCCTTTATTTTTCCCCCCCTACAGTTTCCATTT
TTTTTTTATACTTGCTTACTCCAGCCATCTGCAGTACACCAGTTTCAGGTCTTTTGAGCTGTGTAGAGTTTCTGTGT
GTACAGATGTGTGCTCGGACTTTTCTCTTTTTGAGAAATCTGAAGGAGATGGTTGCAGAAGATCCACTTACTACTGA
GAACCATTACCACCGACTCGGCCTCCGGGGTGTTGGGTGGTTTCTGGGTGGTTCCTGGAGCCTCCTCTGGGCAGTGC
ACTGTCCCATCTGTACGCCCTAATGTGCCATTCCCTAGAGGGGAACAACCAAGTGCCGTGGAGGCAGATGATCATGG
TCTGCCTCAACTGTCTGGTTTCCTGTAAAATAAACACATTGTTTTATATTTTTAGGGAACAAAAAGTGCTGCTATAG
GGTTCAAAGTTTTCCTTCTGAACACTTTTCCGAAACAAATTACCCCAAAGACACATTTTGAATATCCTGGTCACATC
TTTGGATCTGTAAAATATACCTTTTAGTATGGCACCTGTTAAAATGCAAAGCAAATTTCTTTGGGGCAGAAAAACAA
TCTGACAGTAGCAGTGTAGAATTTGTTCATTCAAATACATCTGTGTAAATGCAAAAAGTCATAAAATTCACCTCCGA
GCTGCTTGCTTTTGAACCTGCAGCAACTAGTCTTAGCCGGCCCGGTTTGAACATCGTTCTTTCAGAAGTGCTGAAAA
TGCTGCAAAGTTGGATAAGTGGAAATGTGGCTGCCCCTCTCCTCACTACTTCCTCTCTGATCGTTCTGAAGCTTGCA
TTGGGAATGGCTGCTTTCTCTAACCATTTTCAGCTTGAGTGGGTATTGCTGAAGAAATCCAACATCATTCCAGCAGT
TGAAAAAGGAAGCCTTCGGGAGAAAGTGCTTGTCAAAATTTTGTTCTTTGTGCTTGTGTATGAGTAAGTTGCCATGA
ATAAGTTATTATTTTAACCCATAATTGGCGACTGTTTATATGAATTCTTTCTTTGGCACCAAATAGGTTTCATCTTC
TTAGGCACAATTAGAAAAAATCCACATAGATGGATATTTTACATTTAGTTATTGCTTTATCCAAATACATGAATCTA
AAGCTGAATCAACCCTTACTTCCAGTTGTGCTTATTAAGAAGATCAATTTCCAAGTAGTAAAGTTTTCAGGGAAACT
GACTGTGCTGCTATTTGTTTTGACAAATTTGGGGGTAAGTCAATGACAACCAAACCAATCTCGGTGGAAACTCCTAT
CCTATCATGTT

<210> SEQ ID NO 635

<211> Length : 933

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 635

>M62096_PEA_1_node_69

GTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTAAAGTGCTAAGAACTGTGCATTGACATCCAAACATTTCT
TGTACAAAATTTCCCTAGCAAAGCAAACCTGCTTTGACTTAATTTATTTGTTAAATGTTGCACTTTGTTTATGTATG
TTTTGTTTTTGGTGGGGAATAAGGAGAGAGAGGACGACAAATTCTATTGAAGTATTTATTTTGTGAAGATGGCAATT
TTGCATTTGTTTAAATTTTTTTCATTCTTTAATTTTGTTATCAGTGCCAGCCCAATATACCTGCTCTACCATTATTT
GCGGTCTGATAAAAGGGTCCTTGTGGGGCAGGTTTTGCAAAGCTTATCAGGTAATAACATATGCCACATAACCTTGT
TGATATGTTTGCTTCTGATTTGGGAAGCTAAACATTGGTGTTTGAGAGGATTGCCAATTATTAATTGTCATTACCAC
TACTCTCCATTACTTTTTGTTTGGAAATTGAACAAAGGTCAGTAATGGTTTTTGGCTCTTGTTAATATCCATCATAA
AATAGATTGTTTTAGATTCTTTCCAGGGTGATTTTTCCCTGGGTACCCCGTTTCTACTTCTAAAGAATTGCTTGGCA
CTTTCATGTTTCAAAGGGAAACATTCGCTTGTAGTTCCATTTTACTTGATCTCTACAAGGGACTGACAACATTTGCT
TTACTTTTATTCACAGAGAAAGTTGGCTTTGATGTCTCTTAAAGATAATTCTGCTAGTTGCTGATCAGCCAGTCAGT
TCACCTAGCTTCAATCTTTATAGGACTTCTAATCTAATTTTCCTATAGTGTGACTAAAAGGGAGGCAAATTATTGGA
ACGGATTATTCAAATGGATCCTTAAATATTGCTATGTATAATAAGCCAGTTATTATATCAGGACCATGTTCTCTGTA
GGCCACTTT

<210> SEQ ID NO 636

<211> Length : 1,247

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 636

>M62096_PEA_1_node_71

TTTAAATGTAAAAACCACACTTCTGAACAACTAAGCTCATGAATATGATTTTGGTTATATGCAGCTTTTGACTAGCA
TGTATTGTGTCTTTTTCTCCTCTATGAATAATTTTATATTTCATGCTACTTCTTGAAAGTTTACTCTTTGATGCTCT
AAGAGAACAGCCAGATGGTTTATATGAATAATCTTTATCTGCAGGATGGTGGATTGGTAAATTAGGAGAATGTTGTT
TGAGATATCAAGATTTATGTCTGGGAACTAAAATATATAATGCCAAATGTGTTTTTGTCAATTACTAGAGAATTCTG
TGCAAACATATCATCTCTTCAAATGCTGCACACTTTGCTTTTGTTAAACAGCAGGTAGTAGACAGAACAATAACAGT
TTCGCGTTAAGACTTTTAAAGGAAATAGAATCGTGATTAAGAAATCAGAATTTATAGATATATTGGGATAAATGAAG

AAATAAAAATGTTTGTCTAGAATGTAGCATCTAGTGACTTTTTAAAGCCCTAACGTTTACATAAAGAAGCTCTAGTT
CTTATAGAAATAACAAAGCAAATAAAAGTTCTTAACAATCCCCTCTTTCGAAGTGCATTTTTTTAAAGCAGGGCAGG
AGACATTTGGACTCTAGCTATATGACATACTGGGAAAGGCAGAGGGTGGAGGGAAGATTTCACTTCATTGTCTAGCC
CAGAATCTTGAGCAAGCTAAAGAAACCATCATAATCTAAAATTGCTTCATTTAACACTAACAATTTAGACTTTTTAA
ACCAAGCATTGAATAATGGCTGGATAACTGCCGAAGTAAGCGCCGCTCCATGAAGTCTGCTTACTTATTTAAAAATT
GTGTATCAGTTTTAAATACTGTTCATTGTGTGCAGATATAAGGGGAATAGGGCATTCTGTAGAATTATACATGTCTA
GTTTGTAAAGTGTGTCCTGTGTACTGCAGATGTGTGTTCTCTGGGCTTTATGTATCTGTACAGTAGCTTTCACATTA
AAAAAATTGTGGACAAACTTGTCCGGGGGGTTTGAGGGGAGAATGGTGGTTTATATCAATAACGATGCTGTACTATA
GTCCATGTAACAAAAGATCTGGAAGTCACCCTCCTCTGGCCCACGGAAAATTTTGGTAATCTTCTAGGTTCTAAAAT
GAAGATGTATGGGTACTCTGGCAGACTGCATGTTGTATAATTTGAAAAATACTAAAAGTGGAAAATAAAATTGAATT
AAACTTTGGCTGGTC

<210> SEQ ID NO 637

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 637
>M62096_PEA_1_node_1
CCCCACCCATCCCCGTGC

<210> SEQ ID NO 638

<211> Length : 91

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 638
>M62096_PEA_1_node_4
CAAGGGAAGCCATATGTCTTCGACAGAGTGCTACCTCCCAACACGACCCAAGAGCAGGTTTACAATGCATGTGCGAA
GCAAATTGTCAAAG

<210> SEQ ID NO 639

<211> Length : 74

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 639
>M62096_PEA_1_node_6
ATGTCCTTGAAGGTTATAACGGGACGATTTTTGCGTATGGGCAGACTTCATCAGGAAAAACCCACACCATGGAG

<210> SEQ ID NO 640

<211> Length : 105

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 640
>M62096_PEA_1_node_7
GGGAAGCTGCATGACCCCCAGCTCATGGGGATCATCCCACGAATTGCCCATGATATCTTTGACCATATCTACTCCAT
GGATGAGAACCTGGAGTTTCACATAAAG

<210> SEQ ID NO 641

<211> Length : 49

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 641
>M62096_PEA_1_node_9
GTTTCCTATTTTGAGATCTACTTGGACAAAATAAGGGACTTACTTGATG

<210> SEQ ID NO 642

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 642
>M62096_PEA_1_node_11
TATCCAAGACCAACTTGGCTGTTCATGAAGATAAAAACAGAGTCCCGTATGTAAAG

<210> SEQ ID NO 643

<211> Length : 88

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 643
>M62096_PEA_1_node_13
GGGTGCACTGAGCGGTTTGTGTCGAGCCCTGAGGAAGTCATGGATGTAATAGATGAAGGCAAAGCAAACCGACACGT
GGCTGTGACAA

<210> SEQ ID NO 644

<211> Length : 105

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 644
>M62096_PEA_1_node_21
GTCAGCAAAACTGGTGCCGAGGGAGCTGTTCTTGACGAAGCTAAAAATATCAATAAGTCTTTGTCTGCTCTTGGAAA
TGTGATCTCTGCTTTGGCAGAAGGGACA

<210> SEQ ID NO 645

<211> Length : 61

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 645
>M62096_PEA_1_node_25
GGTATGAAATCAAGGCTTAGGTGCAAAGCCATTGGATACCATACCTGAGACCACACAGCCA

<210> SEQ ID NO 646

<211> Length : 69

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 646
>M62096_PEA_1_node_33
GATGATGAAATTAACCAGCAGAGCCAGCTGGCTGAAAAGCTGAAGCAACAGATGTTGGATCAGGATGAG

<210> SEQ ID NO 647

<211> Length : 118

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 647
>M62096_PEA_1_node_42

CACGAAGCCAAGATCAAGTCTCTGACAGACTACATGCAGAACATGGAACAGAAGAGGAGGCAGCTAGAAGAGTCCCA
GGACTCGCTCAGCGAAGAGCTGGCAAAGCTCCGAGCCCAGG

<210> SEQ ID NO 648

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 648
>M62096_PEA_1_node_44
AAAAAATGCACGAAGTCAGCTTCCAGGATAAGGAGAAGGAACATCTGACGCGGTTGCAGGATGCTGAAGAAATGAAG

<210> SEQ ID NO 649

<211> Length : 110

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 649
>M62096_PEA_1_node_47
AAGGCGCTGGAGCAGCAGATGGAGAGCCACCGGGAAGCTCACCAGAAGCAGCTGTCCAGACTCCGAGACGAAATTGA
GGAGAAGCAGAAAATCATTGATGAGATTCGGGA

<210> SEQ ID NO 650

<211> Length : 102

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 650
>M62096_PEA_1_node_51
TTTGAATCAGAAACTGCAACTGGAACAGGAGAAGCTTAGTTCTGATTATAACAAGCTGAAAATAGAGGACCAAGAGA
GAGAAATGAAGCTGGAAAAGCTCTT

<210> SEQ ID NO 651
<211> Length : 61
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 651
>M62096_PEA_1_node_53
ATTGCTCAACGATAAAAGGGAACAAGCCAGAGAAGACCTCAAAGGGCTGGAGGAGACAGTG

<210> SEQ ID NO 652
<211> Length : 72
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 652
>M62096_PEA_1_node_55
TCTAGAGAATTGCAGACACTGCACAACCTTCGGAAACTCTTTGTCCAGGATCTGACCACCCGAGTTAAAAAA

<210> SEQ ID NO 653
<211> Length : 105
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 653
>M62096_PEA_1_node_58
AGTGTGGAGTTGGACAACGATGATGGAGGGGGCAGTGCTGCCCAGAAGCAGAAAATTTCCTTCTTGGAGAATAACCT
GGAGCAGCTCACCAAAGTTCACAAGCAG

<210> SEQ ID NO 654
<211> Length : 114
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 654
>M62096_PEA_1_node_62
CCAAGCCCATCCGCCCCGGACACTACCCGGCCTCATCTCCAACGGCCGTCCATGCCATTCGAGGGGGAGGAGGCAGC
TCTTCAAATTCCACTCACTACCAGAAATAAATACAAA

<210> SEQ ID NO 655
<211> Length : 118
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 655
>M62096_PEA_1_node_66
GTGTGCCCAAGATGAGTGAGCTGGCACTGTGCCCTGAAGCTTTCACCACTGTAATGAAATATATGCCAGGGGAGACT
TTGGGCTTTTCTCATGACTGTGTGGGTCGAAGGTAGCTCAA

<210> SEQ ID NO 656
<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 656
>M62096_PEA_1_node_67
GTGTGT

<210> SEQ ID NO 657

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 657
>M62096_PEA_1_node_68
GTGTGT

<210> SEQ ID NO 658

<211> Length : 55

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 658
>M62096_PEA_1_node_70
CTAAAAAAGCCACATATGTGCAATTTTCAGGTTTTTAGACTATTGCTCCCTGTAC

<210> SEQ ID NO 659

<211> Length : 160

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 659
>M78076_PEA_1_node_0
CGCGGGGCGGGGCTGGCGGCGCCGGCGCAGCCCGGGGGCGGCGGGAGGAGGAGGTGGCGGCGGTGGCGCTGGGAGCT
CCTGTCACCGCTGGGGCCGGGCCGGGCGGGAGTGCAGGGGACGTGAGGGCGCAAGGGCCGGGACATGGGGCCCGCCA
GCCCCG

<210> SEQ ID NO 660

<211> Length : 133

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 660
>M78076_PEA_1_node_10
ATGTACCCGGAGCTGCAGATTGCACGTGTGGAGCAGGCTACGCAGGCCATCCCCATGGAGCGCTGGTGCGGGGGTTC
CCGGAGCGGCAGCTGCGCCCACCCCCACCACCAGGTTGTGCCCTTCCGCTGCCTGC

<210> SEQ ID NO 661

<211> Length : 134

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 661
>M78076_PEA_1_node_15
GCCTGCAGCTCCCAGGGCCTCATCCTGCACGGCTCGGGCATGCTCTTACCCTGTGGCTCGGATCGGTTCCGTGGTGT
GGAGTATGTGTGCTGTCCCCCTCCAGGGACCCCCGACCCATCTGGGACAGCAGTTGG

<210> SEQ ID NO 662

<211> Length : 179

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 662
>M78076_PEA_1_node_18
TGACCCCTCCACCCGGTCCTGGCCCCCGGGGAGCAGAGTAGAGGGGGCTGAGGACGAGGAAGAGGAGGAATCCTTCC
CACAGCCAGTAGATGATTACTTCGTGGAGCCTCCGCAGGCTGAAGAGGAAGAGGAAACGGTCCCACCCCCAAGCTCC
CATACACTTGCAGTGGTCGGCAAAG

<210> SEQ ID NO 663

<211> Length : 131

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 663
>M78076_PEA_1_node_20
TCACTCCCACCCCGAGGCCCACAGACGGTGTGGATATTTACTTTGGCATGCCTGGGGAAATCAGTGAGCACGAGGGG
TTCCTGAGGGCCAAGATGGACCTGGAGGAGCGTAGGATGCGCCAGATTAATGAG

<210> SEQ ID NO 664

<211> Length : 159

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 664
>M78076_PEA_1_node_24
CACTTCCAGTCCATTCTGCAGACTCTGGAGGAGCAGGTGTCTGGTGAGCGACAGCGCCTGGTGGAAACCCACGCCAC
CCGCGTCATCGCCCTTATCAACGACCAGCGCCGGGCTGCCTTGGAGGGCTTCCTGGCAGCCCTGCAGGCAGATCCGC
CTCAG

<210> SEQ ID NO 665
<211> Length : 129
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 665
>M78076_PEA_1_node_26
GCGGAGCGTGTCCTGTTGGCCCTGCGGCGCTACCTGCGTGCGGAGCAGAAGGAACAGAGGCACACGCTGCGCCACTA
CCAGCATGTGGCCGCCGTGGATCCCGAGAAGGCACAGCAGATGCGCTTCCAG

<210> SEQ ID NO 666
<211> Length : 1,643
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 666
>M78076_PEA_1_node_29
TCACATCCTTCCAGCTCCCAAATGCGCCGCTATTCCTCAGACGCCCGCGCCTCAGGCTCTTCTCTTGTCCCTTAGAC
CCTCTTTCTGTCTCTTGGACCCCTTCCTATCCCCTGAACACCGCTTCTCTGCCCCTTCCCAGTCTCTCAGCTCAGCT
TCCTGACCCTGAAACATGGACCCTCACATGCTGTGTCTTTGACCCCTGCTTCTTGGCCCTTGGATTCCTACTCCCCC
CGCCGTCGATCCTATGTTCTGTCCCTTGGATTTTCACTGCCTTTCCCAGAATCGTCTTTTTTTTTTTTTTTTTTTG
AGACAGGTTCTTGCTCTGTCGCCCAGGCAGGAGAGCAGTGTGCGATCTTGGCTCATTGCAACTTCCACCTCCTGGGT
TCAAGCAATTCTCCTGCCTCAGCCTCTCGAGTAGCTGGGATTACAGGAGCCTGCCACCACACTGGGCTAATTTTTTT
TTTTTTTTTTGACAGAGTCTCGCTCTGTTTCCCAGGCTGGAGTGCAGTGACATGATCTGGGCTCACTGCAACCTCCG
CCTACTGGGTTCAAGCTATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCGGGTGTCACCACATCTGGC
TGATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATACTGGTCAGGCTGGTCTTGAACTCGACCTCAGGTGATCC

ACCCTTGGCCTCCTAAAGTACTCGGATTACAGGTGTGAGCCACCACGCCCGGCCCCAGCTAATTTTTGTATTTTTGG
TAGACACGGGTTTCAGCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCTGCCTGCCTTGGCCTCCC
AAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCAGCCAGAAACCCCAATAACTTTTGCACCAATCTAATATTTT
TAGCAGAGACAGGGTTTTGCCATGTTGCCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCTGCCCACCTCGGCC
TCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCGGCCAGAAACCCCAATAACTTGCACCAATCTAATATT
TTTAGCAGAGACAGGGTTTTGCCATGTTGCCCAGGCTAGTCTCAAACTCCTGACCTCAGGTGATCTGCCTACCTCGG
CCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGTCGAGAATCTCCTTCTTGTTCCTTGAACCCTCT
TCCTGTCCCTCAACCTCCTTTCTCCATAACTTCACTTGTTTTCCCTGGAACCCCTGTTCTGTGCGCTCAAATTTGAA
TTCCCCTTTCCTGGATGTTTTCTTCCTGTCTATGAAACTCCATTCTGTGCTCTTGAACTCCAAATCTTGCCTTGAAC
CATGTCATTTCTATATGACCCTCCAATCCTCAATCTCTGTCTCTGGAATCCCCTCAAACCCCACTTTCTGTTCCTTG
GACTTTATTCTTCAATTTCCTTCTCCTATGGCCCAGTTCCTAACCCTTGTACCACACATCCTGTCCATTGCATGTGC
CGCTTTTCCTCAGTCGCTATTGAATTCCTCCTTCATACTGCTTCAGTTTCCTCATCTCCAGCCTGCATTGCGCAGTT
CATCCTTCATGTCCACTCACCCACAG

<210> SEQ ID NO 667

<211> Length : 872

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 667

>M78076_PEA_1_node_32

GTGAGTGTCTATTACCCTGGCTCCCATTACAGATCTCTGAGGGCAGATCTTGACTCCTAAATGTTGGGCCCCCCCAA
TTTCATTTATTCCTCTATAACAAACAGCCCAGACCTTAGCAGTGAAAATCAACAATGATTTTTCTTTGTTCATGATT
CTGCCATCCGGTCTGCGCTCAGCAGAGTGGTTCTTTCAGTGGTCTTGCCAGTGGTCAAGCATGCAGCTGTATTTAGC
TAGCAGATCATCTAGGGGCTGGGAGTCTAGCACAAATGGACCTTTCTCTCTCTCCAAGGAAGCGCAAGGCCTCTCTT
CTCCGTGGAGCTTCTCCATGTGGTCTCATCAGCAGGGTAGCTAGATTCCCTACATGGTGGTTTATGCTCTCTAAGAC
ATCACAGTGGAAGTTGCTAGGTCTTAAGGCTTGGGCCCACATTCTATTTGTTAAAGCAAGTTACAAATTCAGTCCAG
ATTCAAGGGAAGGAACCTATATGCATACCGGAAAGTGTGACCTATTGCAGCCCCCACATCTATTGTGTCTTTCTCCT
GGATATCTCACACATAACCCTGATTCTCCTAGTATTTAAGAAAGCTATCATCTTGAGGCGCGGTGGCTCACGCCTAT
AATCCCAGCACTTTAGGAGGCCGAGGCGGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGT
GAAACCCCGTCTTTACTAAAAATACAAAAATCAGCCGGGCATGATGTCGCTTGCCTGTAATCCCAGCTACTTAGGAG
GCTGAGGCAAGAGAATTGCTTGAACCCGGGAGGTGGAGGTTGCAGTGAGCTGAGATCGCATCATTGCACTCCAGCTG
GGCAACAAGAGTGAGACTCTGTCTC

<210> SEQ ID NO 668

<211> Length : 259

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 668

>M78076_PEA_1_node_35

GTGAGTGAGCCCACATATAGATGACCCCAGACATTAGGGAACAGGCCCCAGCCTAATTTGTAATCCCCTAGAGTCTG
AGGGTGTCTTCACCACCACAGTGACTGGGAGAGGATGAGGAGGAACGTCTAAGGTTGCAGGGGCCTCTGTAGGATCC
CCAATCCTCCTTCTTAGTCCCTGGAAGGATGTTTCTCCACCTTTCTTTGCTGATACCCTCCTCTCTTCACTGTTCCA
CTCCCTTGCTTCCTCTGGCTGCCAGCAG

<210> SEQ ID NO 669

<211> Length : 463

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 669

>M78076_PEA_1_node_37

GTGAGTGTCTCACAGTTAACCCCAGCCTCCAAATCCCACTGAATCCCTGAACCCAGAAGGAAACAGGGTCCATCCAT
TGGGAACCTCAGACCCCCTGGGGTAGAGTTTGATGTACTTTCCAGCCCCCTCCTCTGGACCCTAAAGAATGAGATAG
GGCCAGGCGCTGGTGACTCACACCCGTAATCCTAGCACTTTCAGAGGCTGAGGCAGGAGGATCCCTTGAGGCCACGA
GTTCTAGACCAGCCTGGGCAACATAATGAGACCCTGTACCTACAAATAATTTAAAAATTACCTGGGTGTGGTGGGGC
ATGTCTGTAGTCCCAGCTGCTCAGGAGGCTGACGTAGAAGGATCACTGGAGCCCAGGAAGTTGAGGCTGCAGTGAGC
TGAGATCATGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTCTGTCTAAAGAAAAAAAAAAAGAATGAGATCA
G

<210> SEQ ID NO 670

<211> Length : 121

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 670
>M78076_PEA_1_node_46
GTAAGAGGAGGAACAGCCGGGTACCTAGGGGAAGAGACCAGAGGTCAGCGGCCAGGCTGTGATTCCCAAAGCCACAC
AGGACCCTCAAAGAAGCCCTCTGCCCCATCTCCTCTCCCTGCAG

<210> SEQ ID NO 671

<211> Length : 144

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 671
>M78076_PEA_1_node_47
GCACCAGCTGGGACAGGGGTGTCCCGTGAGGCTGTGTCGGGTCTGCTGATCATGGGAGCGGGCGGAGGCTCCCTCAT
CGTCCTCTCCATGCTGCTCCTGCGCAGGAAGAAGCCCTACGGGGCTATCAGCCATGGCGTGGTGGAG

<210> SEQ ID NO 672

<211> Length : 304

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 672
>M78076_PEA_1_node_54
AACCCCAACTCCCAGCCTAGGGCAGCAGGGAGTCTTGAAGTGATCATTTCACACCCTTTTGTGAGACGGCTGGAAAT
TCTTATTTCCCCTTTCCAATTCCAAAATTCCATCCCTAAGAATTCCCAGATAGTCCCAGCAGCCTCCCCACGTGGCA
CCTCCTCACCTTAATTTATTTTTTAAGTTTATTTATGGCTCTTTAAGGTGACCGCCACCTTGGTCCTAGTGTCTATT
CCCTGGAATTCACCCTCTCATGTTTCCCTACTAACATCCCAATAAAGTCCTCTTCCCTACCAGGCCAGTCTGA

<210> SEQ ID NO 673

<211> Length : 44

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 673
>M78076_PEA_1_node_1
CTGCTCGCGGTCTAAGTCGCCGCCCGGGCCAGCCGCCGCTGCCG

<210> SEQ ID NO 674

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 674
>M78076_PEA_1_node_2
CTGCTGCTGCCACTATTGCT

<210> SEQ ID NO 675

<211> Length : 64

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 675
>M78076_PEA_1_node_3

GCTGCTTCTGCGCGCGCAGCCCGCCATCGGGAGCCTGGCCGGTGGGAGCCCCGGCGCGGCCGAG

<210> SEQ ID NO 676

<211> Length : 82

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 676
>M78076_PEA_1_node_6
GCCCCGGGGTCGGCCCAGGTGGCTGGACTATGCGGGCGCCTAACCCTTCACCGGGACCTGCGCACCGGCCGCTGGGA
ACCAG

<210> SEQ ID NO 677

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 677
>M78076_PEA_1_node_7
ACCCACAGCGCTCTCGACGCTGTCTCCGGGACCCGCAGCGCGTGCTGGAGTACTGCAGACAG

<210> SEQ ID NO 678

<211> Length : 113

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 678

>M78076_PEA_1_node_12
CTGGTGAATTTGTGAGTGAGGCCCTGCTGGTGCCTGAAGGCTGCCGGTTCTTGCACCAGGAGCGCATGGACCAATGT
GAGAGTTCAACCCGGAGGCATCAGGAGGCACAGGAG

<210> SEQ ID NO 679

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 679
>M78076_PEA_1_node_22
GTGATGCGTGAATGGGCCATGGCAGACAACCAGTCCAAGAACCTGCCTAAAGCCGACAGACAGGCCCTGAATGAG

<210> SEQ ID NO 680

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 680
>M78076_PEA_1_node_27
GTGC

<210> SEQ ID NO 681

<211> Length : 72

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 681
>M78076_PEA_1_node_30
GTGCATACCCACCTTCAAGTGATTGAGGAGAGGGTGAATCAGAGCCTGGGCCTGCTTGACCAGAACCCCCAC

<210> SEQ ID NO 682

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 682
>M78076_PEA_1_node_31
CTGGCTCAGGAGCTGCGGCCCCAAATCC

<210> SEQ ID NO 683

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 683
>M78076_PEA_1_node_34
AGGAACTCCTCCACTCTGAACACCTGGGTCCCAGTGAATTGGAAGCCCCTGCCCCTGGGGGCAGCAGCGAGGACAAG
GGTGGGCTGCAGCCTCCAGATTCCAAGGATG

<210> SEQ ID NO 684

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 684
>M78076_PEA_1_node_36
ACACCCCCATGACCCTTCCAAAAG

<210> SEQ ID NO 685
<211> Length : 12
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 685
>M78076_PEA_1_node_41
ACTTGGGGGTAG

<210> SEQ ID NO 686
<211> Length : 9
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 686
>M78076_PEA_1_node_42
GGTCCACAG

<210> SEQ ID NO 687
<211> Length : 62
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 687
>M78076_PEA_1_node_43

AACAAGATGCTGCATCCCCTGAGAAAGAGAAGATGAACCCGCTGGAACAGTATGAGCGAAAG

<210> SEQ ID NO 688

<211> Length : 63

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 688
>M78076_PEA_1_node_45
GTGAATGCGTCTGTTCCAAGGGGTTTCCCTTTCCACTCATCGGAGATTCAGAGGGATGAGCTG

<210> SEQ ID NO 689

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 689
>M78076_PEA_1_node_49
GTGGACCCCATGCTGACCCTGGAGGAGCAGCAGCTC

<210> SEQ ID NO 690

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 690
>M78076_PEA_1_node_50
CGCGAACTGCAGCGGCACGGCTATGAGAACCCCACTTAC

<210> SEQ ID NO 691

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 691
>M78076_PEA_1_node_51
CGCTTCCTGGAGGAACGACCCTGACCCGGCCCCCTTCACCCCTTCAGCCGAGCCCAGAC

<210> SEQ ID NO 692

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 692
>M78076_PEA_1_node_52
CTCCCCTCTTCCTGGAG

<210> SEQ ID NO 693

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 694
>M78076_PEA_1_node_53
CCCCAG

<210> SEQ ID NO 695

<211> Length : 307

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 695

>T99080_PEA_4_node_1

GCGGTCAGCCCAAGGTCACTTGACCCAGTCAGTGTCCGGCCAACTCTGCAGCTCGGTCCAGCCCTGCCCTTGGGGAG
CCGGGGAGGGGCGGGAGAGGCTTTTCTGGAGCTCCTTCAAAGAAGAACTTGTACTTTTCTGAGAACGACGCTCCCAG
ACCTTGGGGTGTGCCCTTGTCTGGCAAAGGGCGGAGGCCCTGGCTGTGCCTCCGCGTGCTTCCGCCGCAGGATGCCG
GCGTCCGCCCGCCTGGCGGGAGCGGGGCTGCTGCTGGCCTTTCTCCGCGCGCTCGGCTGCGCTGGGCGGGCCCCAG

<210> SEQ ID NO 696

<211> Length : 447

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 696

>T99080_PEA_4_node_6

GTATGTGGCCTGCAGGCTTTGGGGTGGTAACCTAGGTTGTGGGCATAGGAACAAGGGACGTGTTCTTGACAAATGTG
AACTAGAAGCCGCTGGCTATTTGGTAGCTGCATGGCAAAGAGGTAGTTTGTAGAGCAATGAGTATGAAAATGCTGTG
CAATCGGTAAAACATGGTGTAAATGGAAGATCACCCTGCTGTTATTATTAGTCAGTGGTGCCTGAGCATCTAGATAC
ACATATTAATGAGCTTCTCTCTTCCAAGGGAAATAGAGGGCTTCCCCAGTGCTCGCCGTTGTGGCATCATCAAACCA
AGTCAGGTTTCTTATAGAAAGGCTAACATTGATTGAAGAGCTGGCATTTAGATGACCTGATGTCATGTATATAACAT
ATATAAATGCTTCCTCTGCAGCTGCTGCACTTTCTTCAGACCTCTTTCTCCAAGCTTCCCTC

<210> SEQ ID NO 697

<211> Length : 523

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 697

>T99080_PEA_4_node_11

GTATTAATGAATTTAAAAGGATTTAAATCAAGGAATGTTCTCCAACTACAGTGGAACTAAACCACATTAAAAAAATA
AAAAGGATAACTGGAAAATCCCAAAATATTTGGAAACCATATAGCACACTTACTTCTAAAATTGTGGTAGAATACAT
ATAACATAGAAATTATTGTTCTAACCATTTTTAAATGTACAATTCAGTGGTCTTAAGCACATTCACATTGTTCTGTT
TATCTACAGAACGCTTTTCATCTTGCAAAACTGAAACTCTGTATTCATTAAACACTAACTCCCCATTTTCTCCTTCC
CCCATGCCCCTGACAATCATAAATCTACATTCTATTAATTCAACTGCTCTAGTTACCTCATATAAGTGGAATTTTAC
AGTATTTGTCCTTTTGTGGCTGGCTTATTTCACTTAGCATAATATCCTCAGGGTTCATCTGTGTTATATCATGAAAG
TAAAAACAATTTCCTTTCTTTGTAAGACGGAATAATATCCTGTTGTATGTGTATACTTTCA

<210> SEQ ID NO 698

<211> Length : 1,288

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 698

>T99080_PEA_4_node_19

CTGGGTTAATTTGAAGCAGAAGAGGGCAGTTGTTATACTGCCCTGTCAGTTGGATGCGGAGTCTTACTCAAAATTCA
TTCTCAGCATTCTTCTTTTATGGTATCTTCTTTGGCACTTAGCAGCGCATCAGGTAGGCATCTTCTATTTTTCTTCA
TTCCTTAATTTCCTTTGTATCCCTCAAATGGTTATTTATTTGGCTGGAGTCTGTTTTGTTCATTAAGCAAACATGTC
TTTGCTTTGAACATGTCTTTGATTTGATGGATACTTAAATTCCTCATCAAACATTTTGTTGCTATGCATAACGTTTT
CTTTGGCCAACTCCAGCAATTTCCCACATTTTGACATGCAATCATGTTAACTCCCATTTTCTTTTGTAATCCAACAT
CTTCTATTTAGATAATTACTTTAACAATCAATGACTTAATATTCTAATCATAAATTTATACAAAAATAAAATTACCT
CCAAAACATTGCTACCTTTCCTAAACATTCAGTCTTGCCACAGTTTAATAAAAGGAAAGAACATTAAAAAGGATAAG
ACACTGTAATGATTAGATGCTTTTTATAAGCCTAAAGGCATTGTGATTATTTAGACAGAAGAGAAGAAAGTGAAGTG
AAAACCTGATAGTTATGTAGTCTCATGGTTTGCTGTTGAGAGGCTGAACACCAGCTGCTTTCCTTTTCTAGGAAGAT
AATAAAGTGGGCTTTGGCTACAACATAAAGATGTTGGGTTAGACAGTTTCACTACAGTAAGAACAACGGGATGAGTT
GCCCAGGAAATTGTGAAATACTTTCTAATGATCTTTAAAGATATAATGAACACTAATTCATCTGGATTTGTTTAGGT
GTGGTCCTGGTTAAAGGCAAAGGGAAGGATCAGATAACTTCATGTTTTTTCCATTTAACATACCCAATAGATTCTTG
ATTAGGGGAAGGGAAAATGAGCAAGATACAGTCCAGTATTCTAAAAACAATCAGCCTTAGGGGATCATTTCAAAAGC
ATCTGTTTTGGACTTAAGTCTTTGATACTTAACCAAATTGACTACACAGTGAAAAATTCTAGTGCCTGGGTTTTATA

GGGTAGAAGAAAGACATGCAGTCAAGTGGCCAATACTTCATGTGAAGATAAGCAATGAGATCCTTCTTGCTGTCTTT
CTTTTGACTGTTCTGGGCAATATCAAATTAGTTTCAGTGGCTTGATTCTAGGCCAAGATTCTGGCAACAGATTGTAG
TCTTACCTTGTTTTCTTCAATCTCACTGGATCTCTCTCTCTTTTTACCCCCCTTAG

<210> SEQ ID NO 699

<211> Length : 439

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 699
>T99080_PEA_4_node_20
GCTGAGGGTAAAAAGCTGGGATTGGTAGGCTGGGTCCAGAACACTGACCGGGGCACAGTGCAAGGACAATTGCAAGG
TCCCATCTCCAAGGTGCGTCATATGCAGGAATGGCTTGAAACAAGAGGAAGTCCTAAATCACACATCGACAAAGCAA
ACTTCAACAATGAAAAAGTCATCTTGAAGTTGGATTACTCAGACTTCCAAATTGTAAAATAATGGCCTGAATTTAAG
TTTTCTAAGATAAACTCAGTGGTTTGGTTTTTATTATTAATAGAGATAGAACTATTGTGTGTTAATATTAGCATTAG
TCAATAAGTTATTTTAATGTCAGATTTTTGAATGTTATTATATATTACCTGTATGATGGAAGGATTACCACTGTACA
CAAATCTAATCAATAAAAACGTTAGAACCTTCTGCTTAGAGTACATTTAAAAAA

<210> SEQ ID NO 700

<211> Length : 88

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 700
>T99080_PEA_4_node_3
TCCGGGCGCGGAGGTTTGCGCGCCTTGGTGAGCCGTTGGCGTGGTGGTCCCGGAGTGATCCTGGCAGCCGGTGGGGA
AGACAAGGAGG

<210> SEQ ID NO 701

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 701
>T99080_PEA_4_node_5
GTTTGAGCATGGCAGAAGGAAACACCCTGATATCAGTGGATTATGAAATTTTTGGGAAGGTGCAAGGGGTGTTTTTC
CGTAAGCATACTCAG

<210> SEQ ID NO 702

<211> Length : 79

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 702
>T99080_PEA_4_node_8
GAAATGACTGTTGAAAACAGAATTGCTGAAACTCACAGCAAGAGCTGTGTTCCAGTTAGCTTTGCTACCAGTTATGC
AG

<210> SEQ ID NO 703

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 703
>T99080_PEA_4_node_13
TTTTGAGACAGAGTCTTGCTCTGTTGCCTAGGCTAAAGTGCAGTAGTGCGATCTCGGCTCACTGCAACCTCCACTTC

<210> SEQ ID NO 704

<211> Length : 114

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 704

>T99080_PEA_4_node_15

CTTTGTTTTTGGGAGGCTGAGGAAGGAGGATCATTTGAGCCTGGGAGGTTAAGGCTGCAATAAGCTGTGACTGTGCC

ACCATCCTTCAGAAAAAAAAAAGAAAAAGGAAAAGAG

<210> SEQ ID NO 705

<211> Length : 49

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 705

>T99080_PEA_4_node_18

CTATTTACCATATAAGATAACTCTTAAGAACTGGAGATAGTCAGCTCCC

<210> SEQ ID NO 706

<211> Length : 287

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 706

>T08446_PEA_1_node_2

GCTACGGAAGGGTTTTCAGCAGGGAGAGATGGGACCAGCATGCTCTCCATCCTGTTGCCCCGCCTCACGTCTGGCCC

CTGCTTCTCCAGTCCCCCACCCCAGACCACACACGAAGAAGCAGTCCTGTCCTCAGCCCAGCCCTCACCTCCCCCGA

CCTGCCATCCTGCTTCATGCTCAGGGCGGTGTGTGGAGCGCCCGGGGCTCTGGACCCGCGCTGCCAGATAACAATGC

TCTCGTTGTCTCTTTGCTCCCATCTCTGGGGGCCTCTGATTCTTTCTGCTCTACAG

<210> SEQ ID NO 707

<211> Length : 138

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 707
>T08446_PEA_1_node_9
GGCCGTTCCTGGCCGGTTCTCCGGAGTTACGATGACTTTCGTTCCCTGGATGCCCACCTCCACCGGTGCATATTTGA
CCGGAGGTTCTCCTGCCTTCCGGAGCTTCCCCCGCCCCCCGAGGGTGCCAGGGCTGCCCAG

<210> SEQ ID NO 708

<211> Length : 140

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 708
>T08446_PEA_1_node_15
GCGTGTGCCACCACGCCCAGCACAGAGAGATTTTAGATGGCAACCGTGTGGCATCTGCTGTGGAGGATGAAGGTGCA
GAGGTGGATGGGGAAGCCTTCAGGTGGGGAAGCCTTTGGGTGGGAGAGTCCTGGGACATGTGA

<210> SEQ ID NO 709

<211> Length : 123

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 709
>T08446_PEA_1_node_17

CTGGACAATCACGGCCGGCGACTGCTCCTCAGTGAGGAGGCGTCACTCAATATCCCTGCAGTGGCGGCCGCCCATGT
GATCAAACGGTATACAGCCCAGGCGCCAGATGAGCTGTCCTTTGAG

<210> SEQ ID NO 710

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 710
>T08446_PEA_1_node_25
CTGTGCCACGGCCTCGTGGGAAGCTGGCCGGCCTGCTCCGCACCTTCATGCGCTCCCGCCCTTCTCGGCAGCGGCTG
CGGCAGCGGGGAATCCTGCGACAGAGGGTGTTTGGCTGCGATCTTGGCGAGCACCTCAGCAACTCAGGCCAGGATG

<210> SEQ ID NO 711

<211> Length : 145

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 711
>T08446_PEA_1_node_29
GCACGAGTTTGACAGTGAGAGGATCCCGGAGCTGTCTGGCCCTGCATTCCTGCAGGACATCCACAGCGTGTCCTCCC
TCTGCAAGCTCTACTTCCGAGAGCTTCCGAACCCTCTGCTCACCTACCAGCTCTATGGGAAGTTCAGT

<210> SEQ ID NO 712

<211> Length : 146

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 712
>T08446_PEA_1_node_38
GTCCATGGAGCTGGAGTCAGTGGGAATGGGTGGCGCGGCGGCGTTCCGGGAAGTTCGGGTGCAGTCGGTGGTGGTGG
AGTTTCTGCTCACCCATGTGGACGTCCTGTTCAGCGACACCTTCACCTCCGCCGGCCTCGACCCTGCAG

<210> SEQ ID NO 713
<211> Length : 154
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 713
>T08446_PEA_1_node_43
GCCGCTGCCTGCTCCCCAGGCCCAAGTCCCTTGCGGGCAGCTGCCCCTCCACCCGCCTGCTGACGCTGGAGGAAGCC
CAGGCACGCACCCAGGGCCGGCTGGGGACGCCCACGGAGCCCACAACTCCCAAGGCCCCGGCCTCACCTGCGGAAAG

<210> SEQ ID NO 714
<211> Length : 348
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 714
>T08446_PEA_1_node_51
GCCTCCAGAGGCTGCACAGGCTGCGGCGACCCCACTCCAGCAGCGACGCTTTCCCTGTGGGCCCAGCACCTGCTGGC
TCCTGCGAGAGCCTGTCCTCGTCCTCCTCCTCCGAGTCCTCCTCCTCTGAGTCCTCCTCTTCCTCCTCTGAGTCCTC
AGCAGCTGGGCTGGGGGCACTCTCTGGGTCTCCCTCACACCGTACCTCAGCCTGGCTAGATGATGGTGATGAGCTGG
ACTTCAGCCCACCCCGCTGCCTGGAGGGACTCCGGGGGCTGGACTTTGATCCCTTAACCTTCCGCTGCAGCAGCCCC
ACCCCAGGGGATCCCGCACCTCCCGCCAGCCCAGCACCCC

<210> SEQ ID NO 715
<211> Length : 123

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 715
>T08446_PEA_1_node_52
CCGCCCCTGCCTCTGCCTTCCCACCCAGGGTGACCCCCCAGGCCATCTCGCCCCGGGGGCCCACCAGCCCCGCCTCG
CCTGCTGCCCTAGACATCTCAGAGCCCCTGGCTGTATCAGTGCCAC

<210> SEQ ID NO 716

<211> Length : 177

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 716
>T08446_PEA_1_node_55
AACTGCTGGGGGCTGGGGGAGCACCTGCCTCAGCCACCCCAACACCAGCTCTCAGCCCCGGCCGGAGCCTGCGCCCC
CATCTCATACCCCTGCTGCTGCGAGGAGCCGAGGCCCCGCTGACTGACGCCTGCCAGCAGGAGATGTGCAGCAAGCT
CCGGGGAGCCCAGGGCCCACTCG

<210> SEQ ID NO 717

<211> Length : 392

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 717
>T08446_PEA_1_node_57
GTCCTGATATGGAGTCACCACTGCCACCCCCTCCCCTGTCTCTCCTGCGCCCTGGGGGTGCCCCACCCCCGCCCCCT
AAGAACCCAGCACGCCTCATGGCCCTGGCCCTGGCTGAGCGGGCTCAGCAGGTGGCCGAGCAACAGAGCCAGCAGGA
GTGTGGGGGCACCCCACCTGCTTCCCAATCCCCCTTCCACCGCTCGCTGTCTCTGGAGGTGGGCGGGGAGCCCCTGG
GGACCTCAGGGAGTGGGCCACCTCCCAACTCCCTAGCACACCCGGGTGCCTGGGTCCCGGGACCCCCACCCTACTTA

CCAAGGCAACAAAGTGATGGGAGCCTGCTGAGGAGCCAGCGGCCCATGGGGACCTCAAGGAGGGGACTCCGAGGCCC
TGCCCAG

<210> SEQ ID NO 718

<211> Length : 311

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 718
>T08446_PEA_1_node_59
GTTCCTACCCCCGGCTTCTTCTCCCCAGCCCCCAGGGAGTGCCTGCCACCCTTCCTCGGGGTCCCCAAGCCAGGCTT
GTACCCCCTGGGCCCCCCATCCTTCCAGCCCAGTTCCCCAGCCCCAGTCTGGAGGAGCTCTCTGGGCCCCCCTGCAC
CACTCGACAGGGGAGAGAACCTGTACTATGAGATCGGGGCAAGTGAGGGGTCCCCCTATTCTGGCCCCACCCGCTCC
TGGAGTCCCTTTCGCTCCATGCCCCCCGACAGGCTCAATGCCTCCTACGGCATGCTTGGCCAATCACCCCCACTCCA
CAG

<210> SEQ ID NO 719

<211> Length : 206

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 719
>T08446_PEA_1_node_62
CTCTACGTCAACCTAGCTCTAGGGCCCAGGGGTCCCTCACCTGCCTCTTCCTCCTCCTCTTCCCCTCCTGCCCACCC
CCGAAGCCGTTCAGATCCCGGTCCCCCAGTCCCCCGCCTTCCCCAGAAACAACGGGCACCCTGGGGACCCCGTACCC
CTCATAGGGTGCCGGGTCCCTGGGGCCCTCCTGAGCCTCTCCTGCTCTACAG

<210> SEQ ID NO 720

<211> Length : 426

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 720

>T08446_PEA_1_node_63

GGCAGCCCCGCCAGCCTACGGAAGGGGGGGCGAGCTCCACCGAGGGTCCTTGTACAGAAATGGAGGGCAAAGAGGGG

AGGGGGCTGGTCCCCCACCCCCTTACCCCACTCCCAGCTGGTCCCTCCACTCTGAGGGCCAGACCCGAAGCTACTGC

TGAGCACCAGCTGGGAGGGGCCGTCCTTCCTTCCCTTCACCCTCACTGGATCTTGGCCCAACCAAATCCCTTGTTTT

GTATTTTCTTGAACCCCGACCACTACCCCAGGTTTCTAACTTTGTAACTTGCTTCTGATGTGGGTCCCTAACCTATA

ATCTCAGCTTCCCTACCCTGGACTGAAGGGTCTGCCCATCCCCCCACCACCCTCCATCCTGGGGGCCCTCGCACAAA

TCTGGGGTGGGAGGGGCTAGGCTGACCCCATCCTCCTCTCC

<210> SEQ ID NO 721

<211> Length : 98

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 721

>T08446_PEA_1_node_3

GCACGCAGCACTGACAGCCTGGATGGCCCAGGGGAGGGCTCGGTGCAGCCTCTACCCACTGCTGGGGGGCCCAGTGT

GAAGGGGAAGCCTGGGAAGAG

<210> SEQ ID NO 722

<211> Length : 85

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 722

>T08446_PEA_1_node_5

GCTCTCAGCTCCTCGAGGCCCCTTCCCGCGGCTGGCTGACTGCGCCCATTTCCACTACGAGAACGTTGACTTTGGCC

ACATTCAG

<210> SEQ ID NO 723

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 723
>T08446_PEA_1_node_7
CTCCTGCTGTCTCCAGACCGTGAAGGGCCCAGCCTCTCTGGAGAGAATGAGCTGGTGTTCGGGGGTGCAGGTGACCTG
TCAG

<210> SEQ ID NO 724

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 724
>T08446_PEA_1_node_12
ATGCTGGTGCCACTGCTGCTGCAGTACCTGGAGACACTGTCAGGACTGGTGGACAGTAACCTCAACTGCGGGCCTGT
GCTCACCTGGATGGAG

<210> SEQ ID NO 725

<211> Length : 46

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 725
>T08446_PEA_1_node_13
GTGGGCCTGGGCAGGGGGCTTGGAGATTCCGAGTGGGTGAGGGGGT

<210> SEQ ID NO 726

<211> Length : 78

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 726
>T08446_PEA_1_node_19
GTGGGAGACATTGTCTCGGTGATCGACATGCCACCCACAGAGGATCGGAGCTGGTGGCGGGGCAAGCGAGGCTTCCA
G

<210> SEQ ID NO 727

<211> Length : 67

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 727
>T08446_PEA_1_node_21
GTCGGGTTCTTCCCCAGTGAGTGTGTGGAACTCTTCACAGAGCGGCCAGGTCCGGGCCTGAAGGCGG

<210> SEQ ID NO 728

<211> Length : 60

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 728
>T08446_PEA_1_node_23
ATGCCGATGGCCCCCCATGTGGCATCCCGGCTCCCCAGGGTATCTCGTCTCTGACCTCAG

<210> SEQ ID NO 729

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 729
>T08446_PEA_1_node_27
TGCCCCAGGTGCTGCGCTGCTGCTCCGAGTTCATTGAGGCCCACGGGGGTGGTGGATGGGATCTACCGGCTCTCAGGC
GTGTCTTCCAACATCCAGAGGCTTCG

<210> SEQ ID NO 730

<211> Length : 83

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 730
>T08446_PEA_1_node_32
GAGGCCATGTCAGTGCCTGGGGAGGAGGAGCGTCTGGTGCGGGTGCACGATGTCATCCAGCAGCTGCCCCCACCACA
TTACAG

<210> SEQ ID NO 731

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 731
>T08446_PEA_1_node_34
GACCCTGGAGTACCTGCTGAGGCACCTGGCCCGCATGGCGAGACACAGTGCCAACACCAGCATGCATGCCCGCAACC
TGGCCATTGTCTGGGCACCCAACCTGCTACG

<210> SEQ ID NO 732

<211> Length : 89

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 732
>T08446_PEA_1_node_45
GAGGAAAGGGGAGAGAGGGGAGAAGCAGCGGAAGCCAGGGGGCAGCAGCTGGAAGACGTTCTTTGCACTGGGCCGGG
GCCCCAGTGTCC

<210> SEQ ID NO 733

<211> Length : 57

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 733
>T08446_PEA_1_node_46
CTCGAAAGAAGCCCCTGCCCTGGCTGGGGGGCACCCGTGCCCCACCGCAGCCTTCAG

<210> SEQ ID NO 734

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 734
>T08446_PEA_1_node_48
GCAGCAGACCCGACACCGTCACACTGAGATCTGCCAAGAGCGAGGAGTCTCTGTCATCGCAGGCCAGCGGGGCTG

<210> SEQ ID NO 735

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 735
>T08446_PEA_1_node_54
CCGCTGTCCTAG

<210> SEQ ID NO 736

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 736
>T08446_PEA_1_node_58
GTCAGTGCCCAGCTCAGGGCAGGTGGCGGGGGCAGGGATGCGCCAGAGGCAGCAGCCCAGTCCCCATGTTCTGTCCC
CTCACAG

<210> SEQ ID NO 737

<211> Length : 50

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 737
>T08446_PEA_1_node_60
GTCCCCCGACTTCCTGCTCAGCTACCCGCCAGCCCCCTCCTGCTTTCCCC

<210> SEQ ID NO 738

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 738
>T08446_PEA_1_node_61
CTGACCACCTTGGCTACTCAGCCCCCCAGCACCCTGCTCGGCGCCCTACACCGCCTGAGCCC

<210> SEQ ID NO 739

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 739
>T08446_PEA_1_node_64
CTCCAG

<210> SEQ ID NO 740

<211> Length : 52

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 740
>T08446_PEA_1_node_65
GAGCCCCCAGCATGTCCTGACCTGTGCACGGGGATGGGGGGACAACTCCTAC

<210> SEQ ID NO 741
<211> Length : 67
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 741
>T08446_PEA_1_node_66
CCTTCTTTCCCCACATGCCCCACTAAACCATCTGACAACATTAATGAATAAAATGGTGAAAATGTGA

<210> SEQ ID NO 742
<211> Length : 424
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 742
>HUMCA1XIA_node_0
ACACAGTACTCTCAGCTTGTTGGTGGAAGCCCCTCATCTGCCTTCATTCTGAAGGCAGGGCCCGGCAGAGGAAGGAT
CAGAGGGTCGCGGCCGGAGGGTCCCGGCCGGTGGGGCCAACTCAGAGGGAGAGGAAAGGGCTAGAGACACGAAGAAC
GCAAACCATCAAATTTAGAAGAAAAAGCCCTTTGACTTTTTCCCCCTCTCCCTCCCCAATGGCTGTGTAGCAAACAT
CCCTGGCGATACCTTGGAAAGGACGAAGTTGGTCTGCAGTCGCAATTTCGTGGGTTGAGTTCACAGTTGTGAGTGCG
GGGCTCGGAGATGGAGCCGTGGTCCTCTAGGTGGAAAACGAAACGGTGGCTCTGGGATTTCACCGTAACAACCCTCG
CATTGACCTTCCTCTTCCAAGCTAGAGAGGTCAGAGGAG

<210> SEQ ID NO 743

<211> Length : 168

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 743

>HUMCA1XIA_node_2

CTGCTCCAGTTGATGTACTAAAAGCACTAGATTTTCACAATTCTCCAGAGGGAATATCAAAAACAACGGGATTTTGC

ACAAACAGAAAGAATTCTAAAGGCTCAGATACTGCTTACAGAGTTTCAAAGCAAGCACAACTCAGTGCCCCAACAAA

ACAGTTATTTCCAG

<210> SEQ ID NO 744

<211> Length : 214

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 744

>HUMCA1XIA_node_4

GTGGAACTTTCCCAGAAGACTTTTCAATACTATTTACAGTAAAACCAAAAAAAGGAATTCAGTCTTTCCTTTTATCT

ATATATAATGAGCATGGTATTCAGCAAATTGGTGTTGAGGTTGGGAGATCACCTGTTTTTCTGTTTGAAGACCACAC

TGGAAAACCTGCCCCAGAAGACTATCCCCTCTTCAGAACTGTTAACATCGCTGACGGGAA

<210> SEQ ID NO 745

<211> Length : 163

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 745

>HUMCA1XIA_node_6

GTGGCATCGGGTAGCAATCAGCGTGGAGAAGAAAACTGTGACAATGATTGTTGATTGTAAGAAGAAAACCACGAAAC
CACTTGATAGAAGTGAGAGAGCAATTGTTGATACCAATGGAATCACGGTTTTTGGAACAAGGATTTTGGATGAAGAA
GTTTTTGAG

<210> SEQ ID NO 746

<211> Length : 129

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 746
>HUMCA1XIA_node_8
GGGGACATTCAGCAGTTTTTGATCACAGGTGATCCCAAGGCAGCATATGACTACTGTGAGCATTATAGTCCAGACTG
TGACTCTTCAGCACCCAAGGCTGCTCAAGCTCAGGAACCTCAGATAGATGAG

<210> SEQ ID NO 747

<211> Length : 215

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 747
>HUMCA1XIA_node_9
GTGAGGAGCACAAGACCAGAGAAGGTCTTCGTATTTCAGTGATATGGACATAGCAGTGCAGTTTTTGAACTTATACT
TATTTATTCCATTTTAATTAAGCTATGCTTTGTATTTTAATTGTGTTGTAATATTTCCAGGAAAAAGTGACTTGAAT
ATATTTGGTACTTGTTTTCTTGCTGTTTAAGCATTTGATTACATAAATTTATTAGCAATAA

<210> SEQ ID NO 748

<211> Length : 214

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 748
>HUMCA1XIA_node_18
CCAAATCCAGTTGAAGAAATATTTACTGAAGAATATCTAACGGGAGAGGATTATGATTCCCAGAGGAAAAATTCTGA
GGATACACTATATGAAAACAAAGAAATAGACGGCAGGGATTCTGATCTTCTGGTAGATGGAGATTTAGGCGAATATG
ATTTTTATGAATATAAAGAATATGAAGATAAACCAACAAGCCCCCCTAATGAAGAATTTG

<210> SEQ ID NO 749
<211> Length : 430
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 749
>HUMCA1XIA_node_54
GTCTCCTTTTCTTTCTCATTATTTTACAAAAAAGTAATTAAGTTTGCTTGTGACAAAAGATTTGTAGGAAGACATGA
TGAAAGGAAAGTTGTGAAGTTGTCATTGCCATTATATCTCATATATGAATAGCAAATGATGTTCTTGTTAATAACTC
AATGTCTGGATCAAAACAAGAATAAATCTATAATTAAACACATGTCTTTTTCCCTGATCTCCACTGTGAGATTCCTT
GAGTAATATTTTGTCCCCTGTAGTCATAGCACATTTCTATCTGGCTCTTTCCAACACCTTTTTTCTTTCATTATTTT
TGTTGATTTCTACAAACATATTAATTAAAAAAAACTAATAGCTTTATCGAAGTGTAATTTAAATACTATAAATATTC
CCTTGTTTTAAGTGTACAAGTGAATTATTTTTACTAAATTTACAG

<210> SEQ ID NO 750
<211> Length : 639
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 750
>HUMCA1XIA_node_55
ATGTGCTGCAATCTAAGTTTCGGAATACTTATACCACTCCAGAAATAATCCTCGTGTTAATATTCAGTCCATGTTTC
CACCCCCTCAGACCTAGCTAACAATTAATTTGCTTTCTGTCTCTGTAGATTTGAGTTTTCTGAATATTTCATGTGAA
TGAAACTTATGAATAAATTTATGATTTAATCATATGAATGTGTATGTGACCTTTATGTCTGACTTCTTTCATTTTAG

GTTCATCCATGCTGTAGCATACATATATTAGTACTTTGCTCCTTATTTTGTTCTCATAGTATTCCATTGATTGGGTA
TACCAGGTTCTGTTTACTTTTACTTGGCAGTTGATAGAATAGGTGTAGTTTATACTTTTTCGCTATTCTCCATACCG
GTGCTGTAGTGAATAATTGCATACAAGTCTTTGTATAGATGTGTTTTCATTCTTTTTGGTATATACTTAGAAGCAGA
ATTCTTGTGTTATGGTAAACTTATATTCAATATTTTGTGAATTCCACTCTTTTCCATATCGATTGTACCATTTTCCC
TTCCAAGTAACCATGTATGAGGATAGTCATTTCTGCACATTCTCACTAATGCTTGTTATTGTCTGTCTTCTTGATTA
CGATCATTCTCGTTGGTGTGAAA

<210> SEQ ID NO 751

<211> Length : 129

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 751
>HUMCA1XIA_node_92
GTAAGTATGATGATAATAAATAGCCAGACAATCATGGTTGTGAATTACAGCTCTTCTTTCATTACTCTCATGCTGTG
ATTCCACAGTGTGTGGGAGAGAAAATAAACACATGTCAATCAAATCAGTCAA

<210> SEQ ID NO 752

<211> Length : 117

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 752
>HUMCA1XIA_node_11
TATGCACCAGAGGATATAATCGAATATGACTATGAGTATGGGGAAGCAGAGTATAAAGAGGCTGAAAGTGTAACAGA
GGGACCCACTGTAACTGAGGAGACAATAGCACAGACGGAG

<210> SEQ ID NO 753

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 753
>HUMCA1XIA_node_15
GCAAACATCGTTGATGATTTTCAAGAATACAACTATGGAACAATGGAAAGTTACCAGACAGAAGCTCCTAGGCATGT
TTCTGGGACAAATGAG

<210> SEQ ID NO 754

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 754
>HUMCA1XIA_node_19
GTCCAGGTGTACCAGCAGAAACTGATATTACAGAAACAAGC

<210> SEQ ID NO 755

<211> Length : 63

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 755
>HUMCA1XIA_node_21
ATAAATGGCCATGGTGCATATGGAGAGAAAGGACAGAAAGGAGAACCAGCAGTGGTTGAGCCT

<210> SEQ ID NO 756

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 756
>HUMCA1XIA_node_23
GGTATGCTTGTCGAAGGACCACCAGGACCAGCAGGACCTGCA

<210> SEQ ID NO 757

<211> Length : 63

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 757
>HUMCA1XIA_node_25
GGTATTATGGGTCCTCCAGGTCTACAAGGCCCCACTGGACCCCCTGGTGACCCTGGCGATAGG

<210> SEQ ID NO 758

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 758
>HUMCA1XIA_node_27
GGCCCCCCAGGACGTCCTGGCTTACCAGGGGCTGATGGTCTACCTGGTCCTCCTGGTACTATGTTGATGTTACCG

<210> SEQ ID NO 759

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 759
>HUMCA1XIA_node_29
TTCCGTTATGGTGGTGATGGTTCCAAAGGACCAACCATCTCTGCTCAGGAAGCTCAGGCTCAAGCTATTCTTCAGCA
GGCTCGG

<210> SEQ ID NO 760

<211> Length : 57

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 760
>HUMCA1XIA_node_31
ATTGCTCTGAGAGGCCCACCTGGCCCAATGGGTCTAACTGGAAGACCAGGTCCTGTG

<210> SEQ ID NO 761

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 761
>HUMCA1XIA_node_33
GGGGGGCCTGGTTCATCTGGGGCCAAAGGTGAGAGTGGTGATCCAGGTCCTCAG

<210> SEQ ID NO 762

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 762
>HUMCA1XIA_node_35
GGCCCTCGAGGCGTCCAGGGTCCCCCTGGTCCAACGGGAAAACCTGGAAAAAGG

<210> SEQ ID NO 763
<211> Length : 54
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 763
>HUMCA1XIA_node_37
GGTCGTCCAGGTGCAGATGGAGGAAGAGGAATGCCAGGAGAACCTGGGGCAAAG

<210> SEQ ID NO 764
<211> Length : 54
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 764
>HUMCA1XIA_node_39
GGAGATCGAGGGTTTGATGGACTTCCGGGTCTGCCAGGTGACAAAGGTCACAGG

<210> SEQ ID NO 765
<211> Length : 54
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 765

>HUMCA1XIA_node_41
GGTGAACGAGGTCCTCAAGGTCCTCCAGGTCCTCCTGGTGATGATGGAATGAGG

<210> SEQ ID NO 766

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 766
>HUMCA1XIA_node_43
GGAGAAGATGGAGAAATTGGACCAAGAGGTCTTCCAGGTGAAGCT

<210> SEQ ID NO 767

<211> Length : 0

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 767
>HUMCA1XIA_node_45
GGCCCACGAGGTTTGCTGGGTCCAAGGGGAACTCCAGGAGCTCCAGGGCAGCCT

<210> SEQ ID NO 768

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 768
>HUMCA1XIA_node_47

GGTATGGCAGGTGTAGATGGCCCCCCAGGACCAAAAGGGAACATG

<210> SEQ ID NO 769

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 769

>HUMCA1XIA_node_49

GGTCCCCAAGGGGAGCCTGGGCCTCCAGGTCAACAAGGGAATCCAGGACCTCAG

<210> SEQ ID NO 770

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 770

>HUMCA1XIA_node_51

GGTCTTCCTGGTCCACAAGGTCCAATTGGTCCTCCTGGTGAAAAA

<210> SEQ ID NO 771

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 771

>HUMCA1XIA_node_57

GGACCACAAGGAAAACCAGGACTTGCTGGACTTCCTGGTGCTGATGGGCCTCCT

<210> SEQ ID NO 772

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 772
>HUMCA1XIA_node_59
GGTCATCCTGGGAAAGAAGGCCAGTCTGGAGAAAAGGGGGCTCTG

<210> SEQ ID NO 773

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 773
>HUMCA1XIA_node_62
GGTCCCCCTGGTCCACAAGGTCCTATTGGATACCCGGGCCCCCGGGGAGTAAAG

<210> SEQ ID NO 774

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 774
>HUMCA1XIA_node_64
GGAGCAGATGGTGTCAGAGGTCTCAAGGGATCTAAAGGTGAAAAG

<210> SEQ ID NO 775

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 775
>HUMCA1XIA_node_66
GGTGAAGATGGTTTTCCAGGATTCAAAGGTGACATGGGTCTAAAAGGTGACAGA

<210> SEQ ID NO 776

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 776
>HUMCA1XIA_node_68
GGAGAAGTTGGTCAAATTGGCCCAAGAGGGGAAGATGGCCCTGAAGGACCCAAAGGTCGAGCAGGCCCAACTGGAGA
CCCAGGTCCTTCAGGTCAAGCAGGAGAAAAG

<210> SEQ ID NO 777

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 777
>HUMCA1XIA_node_70
GGAAAACTTGGAGTTCCAGGATTACCAGGATATCCAGGAAGACAAGGTCCAAAG

<210> SEQ ID NO 778

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 778
>HUMCA1XIA_node_72
GGTTCCACTGGATTCCCTGGGTTTCCAGGTGCCAATGGAGAGAAAGGTGCACGG

<210> SEQ ID NO 779

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 779
>HUMCA1XIA_node_74
GGAGTAGCTGGCAAACCAGGCCCTCGGGGTCAGCGTGGTCCAACG

<210> SEQ ID NO 780

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 780
>HUMCA1XIA_node_76
GGTCCTCGAGGTTCAAGAGGTGCAAGAGGTCCCACTGGGAAACCTGGGCCAAAG

SEQ ID NO:781
> AAQ89265
MSLLPRRAPPVSMRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVIITTKSVSR
YRGQEHCLHPKLQSTKRFIKWYNAWNEKRRVYEE

<210> SEQ ID NO 782

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 782
>HUMCA1XIA_node_78
GGCACTTCAGGTGGCGATGGCCCTCCTGGCCCTCCAGGTGAAAGA

<210> SEQ ID NO 783

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 783
>HUMCA1XIA_node_81
GGTCCTCAAGGACCTCAGGGTCCAGTTGGATTCCCTGGACCAAAAGGCCCTCCT

<210> SEQ ID NO 784

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 784
>HUMCA1XIA_node_83
GGACCACCTGGGAAGGATGGGCTGCCAGGACACCCTGGGCAACGTGGGGAGACT

<210> SEQ ID NO 785

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 785
>HUMCA1XIA_node_85
GGATTTCAAGGCAAGACCGGCCCTCCTGGGCCAGGGGGAGTGGTTGGACCACAG

<210> SEQ ID NO 786

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 786
>HUMCA1XIA_node_87
GGACCAACCGGTGAGACTGGTCCAATAGGGGAACGTGGGCATCCTGGCCCTCCTGGCCCTCCTGGTGAGCAAGGTCT
TCCTGGTGCTGCAGGAAAAGAAGGTGCAAAG

<210> SEQ ID NO 787

<211> Length : 90

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 787
>HUMCA1XIA_node_89
GGTGATCCAGGTCCTCAAGGTATCTCAGGGAAAGATGGACCAGCAGGATTACGTGGTTTCCCAGGGGAAAGAGGTCT
TCCTGGAGCTCAG

<210> SEQ ID NO 788

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 788
>HUMCA1XIA_node_91
GGTGCACCTGGACTGAAAGGAGGGGAAGGTCCCCAGGGCCCACCAGGTCCAGTT

<210> SEQ ID NO 789

<211> Length : 211

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 789
>T11628_PEA_1_node_7
GCAAGCCAAAACCCTGGGCAGACTCAATCCAAAAATAAACAATCAAAGAGCATGTTGGCCTGGTCCTTTGCTAGGTA
CTGTAGAGCAGGTGAGAGAGTGAGGGGGAAGGACTCCAAATTAGACCAGTTCTTAGCCATGAAGCAGAGACTCTGAA
GCCAGACTACCTGGGTCCCAATCTTGGGCTTGGTATTTCCTCGCTGTGTGACTCTGG

<210> SEQ ID NO 790

<211> Length : 131

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 790
>T11628_PEA_1_node_11
CCCAGATTTACCAAAGGGAATTGTCAGCTGTCCAAGGGCTAGCAAATTCCTAGGTCACCTAGATTGGATTTTCTGAC
CATAAAAACTGTGGGCCAGGTGCACAGCTGCCTGAGGGGCTCAAACCTGTGCAG

<210> SEQ ID NO 791

<211> Length : 214

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 791
>T11628_PEA_1_node_16
TCCTCCCCTTCTTTCCACACGCACAACCACCCCACCCCCTGTGGCCTGAGCTGTCCTGCCTCGCCACAATGGCACCT
GCCCTAAAATAGCTTCCCATGTGAGGGCTAGAGAAAGGAAAAGATTAGACCCTCCCTGGATGAGAGAGAGAAAGTGA
AGGAGGGCAGGGGAGGGGGACAGCGAGCCATTGAGCGATCTTTGTCAAGCATCCCAGAAG

<210> SEQ ID NO 792

<211> Length : 140

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 792
>T11628_PEA_1_node_22
GCCTTATTTCTCTGCTGGTTGAGCGAAGGGATTGTCTTCCATGGTCTCCGAGATCCCGTCCCACGCTCATGCCCTAG
AATTCTCTGAGTCCTTGATGCACTTTTGCCTTTGGCGAGGAGGCAGGACAGTCAGGCGTGGAG

<210> SEQ ID NO 793

<211> Length : 143

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 793
>T11628_PEA_1_node_25

CTGAGGACTTAAAGAAGCATGGTGCCACCGTGCTCACCGCCCTGGGTGGCATCCTTAAGAAGAAGGGGCATCATGAG
GCAGAGATTAAGCCCCTGGCACAGTCGCATGCCACCAAGCACAAGATCCCCGTGAAGTACCTGGAG

<210> SEQ ID NO 794

<211> Length : 130

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 794
>T11628_PEA_1_node_31
CCCCACCCATCTGGGCCCCGGGTTCAAGAGAGAGCGGGGTCTGATCTCGTGTAGCCATATAGAGTTTGCTTCTGAGT
GTCTGCTTTGTTTAGTAGAGGTGGGCAGGAGGAGCTGAGGGGCTGGGGCTGGG

<210> SEQ ID NO 795

<211> Length : 140

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 795
>T11628_PEA_1_node_37
CTTTGCAATGAGGAGGAGATCTGGGCTGGGCGGGCCAGCTGGGGAAGCATTTGACTATCTGGAACTTGTGTGTGCCT
CCTCAGGTATGGCAGTGACTCACCTGGTTTTAATAAAACAACCTGCAACATCTCAAAAAAAGC

<210> SEQ ID NO 796

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 796
>T11628_PEA_1_node_0
TGTCTCTGAGAGCATTGATGAGGTCAGGAAGCCTCCTGTTGGGTAGAGGAGCAACTAAGAGACTGAACTTGGCCCCC
ACCCTGAGGCTCACAA

<210> SEQ ID NO 797
<211> Length : 103
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 797
>T11628_PEA_1_node_4
GCTTGAATTGCACCTGAGTTCCAAAGGAGAAGTTGACATTCTTCCAGAACATATGCCCAGTGTCTTCAACTTGAGAT
GGAGCTGGGATGCCAAGTCTGCAAAT

<210> SEQ ID NO 798
<211> Length : 47
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 798
>T11628_PEA_1_node_9
CTTGGCTGGAGGCTCTGCGAGGACAGCTGGGGAGAAGGGGAGCTGTG

<210> SEQ ID NO 799
<211> Length : 18
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 799
>T11628_PEA_1_node_13
ATGGAGGCTCGCTCTGTT

<210> SEQ ID NO 800
<211> Length : 98
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 800
>T11628_PEA_1_node_14
GCCAGGCTGGAGTACAGCGATCTCGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCT
CCCAAGTAGCTGGGACTACAG

<210> SEQ ID NO 801
<211> Length : 96
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 801
>T11628_PEA_1_node_17
GTATAAAAACGCCCTTGGGACCAGGCAGCCTCAAACCCCAGCTGTTGGGGCCAGGACACCCAGTGAGCCCATACTTG
CTCTTTTTGTCTTCTTCAG

<210> SEQ ID NO 802
<211> Length : 78
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 802
>T11628_PEA_1_node_18
ACTGCGCCATGGGGCTCAGCGACGGGGAATGGCAGTTGGTGCTGAACGTCTGGGGGGAAGGTGGAGGCTGACATCCCA
G

<210> SEQ ID NO 803

<211> Length : 25

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 803
>T11628_PEA_1_node_19
GCCATGGGCAGGAAGTCCTCATCAG

<210> SEQ ID NO 804

<211> Length : 80

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 804
>T11628_PEA_1_node_24
GCTCTTTAAGGGTCACCCAGAGACTCTGGAGAAGTTTGACAAGTTCAAGCACCTGAAGTCAGAGGACGAGATGAAGG
CGT

<210> SEQ ID NO 805

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 805
>T11628_PEA_1_node_27
TTCATCTCGGAATGCATCATCCAGGTTCTGCAGAGCAAGCATCCCGGGGACTTTGGTGCTGATGCCCAGGGGGCCAT
GAACAAG

<210> SEQ ID NO 806

<211> Length : 29

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 806
>T11628_PEA_1_node_28
GCCCTGGAGCTGTTCCGGAAGGACATGGC

<210> SEQ ID NO 807

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 807
>T11628_PEA_1_node_29
CTCCAACTACAAGGAGCTGGGCTTCCAG

<210> SEQ ID NO 808

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 808

>T11628_PEA_1_node_30
GGCTAGGCCCCTGCCGCTCCCAC

<210> SEQ ID NO 809

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 809
>T11628_PEA_1_node_32
GTGTTGAAGTTGG

<210> SEQ ID NO 810

<211> Length : 22

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 810
>T11628_PEA_1_node_33
CTTTGCATGCCCAGCGATGCGC

<210> SEQ ID NO 811

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 811
>T11628_PEA_1_node_34

CTCCCTGTGGGATGTCATCACCCTGGGAACCGGGAGTGGCCCTTG

<210> SEQ ID NO 812

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 812
>T11628_PEA_1_node_35
GCTCACTGTGTTCTGCATGGTTTGGATCTGAATTAATTGTCCTTTCTTCTAAATCC

<210> SEQ ID NO 813

<211> Length : 118

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 813
>T11628_PEA_1_node_36
CAACCGAACTTCTTCCAACCTCCAAACTGGCTGTAACCCCAAATCCAAGCCATTAACTACACCTGACAGTAGCAATT
GTCTGATTAATCACTGGCCCCTTGAAGACAGCAGAATGTCC

<210> SEQ ID NO 814

<211> Length : 178

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 814
>HUMCEA_PEA_1_node_0

CTCAGGGCAGAGGGAGGAAGGACAGCAGACCAGACAGTCACAGCAGCCTTGACAAAACGTTCCTGGAACTCAAGCTC
TTCTCCACAGAGGAGGACAGAGCAGACAGCAGAGACCATGGAGTCTCCCTCGGCCCCTCCCCACAGATGGTGCATCC
CCTGGCAGAGGCTCCTGCTCACAG

<210> SEQ ID NO 815

<211> Length : 278

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 815
>HUMCEA_PEA_1_node_2
CCTCACTTCTAACCTTCTGGAACCCGCCCACCACTGCCAAGCTCACTATTGAATCCACGCCGTTCAATGTCGCAGAG
GGGAAGGAGGTGCTTCTACTTGTCCACAATCTGCCCCAGCATCTTTTTGGCTACAGCTGGTACAAAGGTGAAAGAGT
GGATGGCAACCGTCAAATTATAGGATATGTAATAGGAACTCAACAAGCTACCCCAGGGCCCGCATACAGTGGTCGAG
AGATAATATACCCCAATGCATCCCTGCTGATCCAGAACATCATCCAG

<210> SEQ ID NO 816

<211> Length : 400

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 816
>HUMCEA_PEA_1_node_11
GTGAGTATATCTGCTCCTCTCTGGCCCAGGCTGCCAGCCCAAATCCACAGGGCCAGAGGCAGGATTTCTCAGTCCCT
CTCAGGTTCAAGTACACAGACCCTCAACCCTGGACATCCAGACTGTCTGTGACTTTCTGCCCCAGAAAAACCTGGGC
AGACCAAGTCTTGACCAAGAATAGGAGGGGAGGGGCTGCTTCTGTCCTGGGAGGCTCAGGGTCCACACCCTATGATG
GGAGAAACAGGTGAATATCTCAGACTCAGGCTCAGTAGATACAAGAGGGGTTTGGCTGAGACTTTAGGATTGTGATT
CAGCTTAGAGGGACACTGTGGTCCTTCCATAGACCAGGAACTTCCACTTCCCTCTGACAATATCACCTGTGGCTTTA
TTTTGTTTGCTCCAG

<210> SEQ ID NO 817

<211> Length : 255

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 817

>HUMCEA_PEA_1_node_12

ATGGCCCGGATGCCCCCACCATTTCCCCTCTAAACACATCTTACAGATCAGGGGAAAATCTGAACCTCTCCTGCCAC
GCAGCCTCTAACCCACCTGCACAGTACTCTTGGTTTGTCAATGGGACTTTCCAGCAATCCACCCAAGAGCTCTTTAT
CCCCAACATCACTGTGAATAATAGTGGATCCTATACGTGCCAAGCCCATAACTCAGACACTGGCCTCAATAGGACCA
CAGTCACGACGATCACAGTCTATG

<210> SEQ ID NO 818

<211> Length : 190

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 818

>HUMCEA_PEA_1_node_31

CTCTCCTGCCATGCAGCCTCTAACCCACCTGCACAGTATTCTTGGCTGATTGATGGGAACATCCAGCAACACACACA
AGAGCTCTTTATCTCCAACATCACTGAGAAGAACAGCGGACTCTATACCTGCCAGGCCAATAACTCAGCCAGTGGCC
ACAGCAGGACTACAGTCAAGACAATCACAGTCTCTG

<210> SEQ ID NO 819

<211> Length : 127

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 819

>HUMCEA_PEA_1_node_36

CTGTCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAATGACGCAAGAGCCTATGTATGTGGAATCCA
GAACTCAGTGAGTGCAAACCGCAGTGACCCAGTCACCCTGGATGTCCTCT

<210> SEQ ID NO 820

<211> Length : 255

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 820
>HUMCEA_PEA_1_node_44
ATGGGCCGGACACCCCCATCATTTCCCCCCCAGACTCGTCTTACCTTTCGGGAGCGAACCTCAACCTCTCCTGCCAC
TCGGCCTCTAACCCATCCCCGCAGTATTCTTGGCGTATCAATGGGATACCGCAGCAACACACACAAGTTCTCTTTAT
CGCCAAAATCACGCCAAATAATAACGGGACCTATGCCTGTTTTGTCTCTAACTTGGCTACTGGCCGCAATAATTCCA
TAGTCAAGAGCATCACAGTCTCTG

<210> SEQ ID NO 821

<211> Length : 1,174

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 821
>HUMCEA_PEA_1_node_46
CTGGGGTGGAGTCTATCTGGTTCTCACCAAAGAGCCAAGAAGACATTTTCTTTCCCAGTCTGTGTTCCATGGGCACA
AGGAAATCCCAAATTCTATCCTGAGCCCCCTCACTCCATCTCGGCCAACTCTCTCCTCCCCGGCTTCTCTGATATCT
CACGGCTGACCTCGGGTCCAGCCTGGAATGTGGGGAGGGGCCTCCCTTAGCCCCAGAAGGCCCCCAATAGTGAAAGG
GACTTCATAGTCCAGAAGAAAGAAGGGTCCTTAAGGTCGAGTTGCTCCTCTCTATCACCAATATGTCCCTTTCTGTC
ACCTCTTTGTGTTTTTTCACCTACTCTGTGAGCTACAAGGAACAAGGAGGCTTTGAAACCAGCCCACACTTTTTCCC
CAAATGAGAGGAGGAAGCCCCTTGGATGAGGCAGGAGCAGCTCAGACTCTGCTCCCTGCTCTGCGCCCGGCTCACCC
GGTGACTGGCTCTGCCCTGGCTCCACTTGGGGTGGGACCGGGGCATGTGGAGAAGGTGTCCAGGTGGCCTGTTTTGA
ATCTGGGTAAATCAAGCTGCCAATCCACAGCAGAGCCTCCCTTGGGTCAGGTTGCAGGGAAATGGGAAAAGAGGGAG
CCTCGGGACAGACTCCTGAGCTGTGTCCTGGCTCTGAAGTCACTGGCTGTATGAGGCTGTGGACACAGCACATAGGA
CACAGCAGAGGAAAGTGAGTGACACACACTTGGAGAAATAGGGAGATTCAGCCATAGGGGCTCTGCATGGGAGGGAA

CAGGCAGTGCCAAAAAGTGTGTGTTTATAGAGAGGGTAAGACTATCAGCCACTATATATATCTAACATAAAACTTAC
CATTAACCATTTCTAAGTGTACAATTAAGTGAAACAGCATAAATATCAATCAAGTATATTGCCCGGTGTGGTGGCTC
ATCCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGAGTGGATCACCTGAGGTCAGGAGTTCAAGATACAGAAAAAA
AAAAATAGCTAGGCATGGTGGTGGGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTCGAACC
TGGGCGGTGTAGTTTGCAGTGAGCCGAGATTGAGCCACTGCACTCCAGCCTGGGTGACAGAGTGAGACTACATCACA
AAAAAAAAAAAAAAAAAGG

<210> SEQ ID NO 822

<211> Length : 179

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 822
>HUMCEA_PEA_1_node_63
CTCAAAAAGAAAAGAAAAGAAGACTCTGACCTGTACTCTTGAATACAAGTTTCTGATACCACTGCACTGTCTGAGAA
TTTCCAAAACTTTAATGAACTAACTGACAGCTTCATGAAACTGTCCACCAAGATCAAGCAGAGAAAATAATTAATTT
CATGGGACTAAATGAACTAATGAGG

<210> SEQ ID NO 823

<211> Length : 732

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 823
>HUMCEA_PEA_1_node_65
TTTGCTGATTCTTTAAATGTCTTGTTTCCCAGATTTCAGGAAACTTTTTTTCTTTTAAGCTATCCACAGCTTACAGC
AATTTGATAAAATATACTTTTGTGAACAAAAATTGAGACATTTACATTTTCTCCCTATGTGGTCGCTCCAGACTTGG
GAAACTATTCATGAATATTTATATTGTATGGTAATATAGTTATTGCACAAGTTCAATAAAAATCTGCTCTTTGTATA
ACAGAATACATTTGAAAACATTGGTTATATTACCAAGACTTTGACTAGAATGTCGTATTTGAGGATATAAACCCATA
GGTAATAAACCCACAGGTACTACAAACAAAGTCTGAAGTCAGCCTTGGTTTGGCTTCCTAGTGTCAATTAAACTTCT
AAAAGTTTAATCTGAGATTCCTTATAAAAACTTCCAGCAAAGCAACTTTAAAAAAGTCTGTGTGGGCCGGGCGCGGT
GGCTCACGCCTGTAATCCCAGCACTTTGATCCGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCCAGACCATCCTG

GCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAGTTAGCCGGGCGTGGTGGTGGGGGCCTGTAGTCC
CAGCTACTCAGGAGGCTGAGGCAGGAGAACGGCATGAACCCGGGAGGCAGGGCTTGCAGTGAGCCAAGATCATGCCG
CTGCACTCCAGCCTGGGAGACAAAGTGAGACTCCGTCAA

<210> SEQ ID NO 824

<211> Length : 280

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 824
>HUMCEA_PEA_1_node_67
CGGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAACCCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGT
GAACCTGAGGTTCAGAACACAACCTACCTGTGGTGGGTAAATGGTCAGAGCCTCCCGGTCAGTCCCAGGCTGCAGCT
GTCCAATGGCAACATGACCCTCACTCTACTCAGCTGTCAAAAGGAACGATGCAGGATCCTATGAATGTGAAATACAG
AACCCAGCGAGTGCCAACCGCAGTGACCCAGTCACCCTGAATGTCCTCT

<210> SEQ ID NO 825

<211> Length : 82

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 825
>HUMCEA_PEA_1_node_3
AATGACACAGGATTCTACACCCTACACGTCATAAAGTCAGATCTTGTGAATGAAGAAGCAACTGGCCAGTTCCGGGT
ATACC

<210> SEQ ID NO 826

<211> Length : 104

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 826
>HUMCEA_PEA_1_node_7
CGGAGCTGCCCAAGCCCTCCATCTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGT
GAACCTGAGACTCAGGACGCAACCTAC

<210> SEQ ID NO 827
<211> Length : 48
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 827
>HUMCEA_PEA_1_node_8
CTGTGGTGGGTAAACAATCAGAGCCTCCCGGTCAGTCCCAGGCTGCAG

<210> SEQ ID NO 828
<211> Length : 48
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 828
>HUMCEA_PEA_1_node_9
CTGTCCAATGGCAACAGGACCCTCACTCTATTCAATGTCACAAGAAAT

<210> SEQ ID NO 829
<211> Length : 79
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 829

>HUMCEA_PEA_1_node_10
GACACAGCAAGCTACAAATGTGAAACCCAGAACCCAGTGAGTGCCAGGCGCAGTGATTCAGTCATCCTGAATGTCCT
CT

<210> SEQ ID NO 830

<211> Length : 3

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 830
>HUMCEA_PEA_1_node_15
CAG

<210> SEQ ID NO 831

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 831
>HUMCEA_PEA_1_node_16
AGCCAC

<210> SEQ ID NO 832

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 832

>HUMCEA_PEA_1_node_17
CCAAACC

<210> SEQ ID NO 833

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 833
>HUMCEA_PEA_1_node_18
CTTCATCACCAGCAACAA

<210> SEQ ID NO 834

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 834
>HUMCEA_PEA_1_node_19
CTCCAACCCCGTGGAGGATGAGGATGCTGTAGCCTTAACCTGTGAACCTGAGATTCAGAACACAACCTAC

<210> SEQ ID NO 835

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 835
>HUMCEA_PEA_1_node_20

CTGTGGTGGGTAAATAATCAGAGC

<210> SEQ ID NO 836

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 836
>HUMCEA_PEA_1_node_21
CTCCCGGTCAGTCCCAGGCTGCAG

<210> SEQ ID NO 837

<211> Length : 78

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 837
>HUMCEA_PEA_1_node_22
CTGTCCAATGACAACAGGACCCTCACTCTACTCAGTGTCACAAGGAATGATGTAGGACCCTATGAGTGTGGAATCCA
G

<210> SEQ ID NO 838

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 838
>HUMCEA_PEA_1_node_23

AACGAATTAAGTGTTGACCACAGCGACCCA

<210> SEQ ID NO 839

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 839
>HUMCEA_PEA_1_node_24
GTCATCCTGAATGTCCTCT

<210> SEQ ID NO 840

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 840
>HUMCEA_PEA_1_node_27
ATGGCCCAGACGACCCCAC

<210> SEQ ID NO 841

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 841
>HUMCEA_PEA_1_node_29
CATTTCCCCCTCATACAC

<210> SEQ ID NO 842

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 842
>HUMCEA_PEA_1_node_30
CTATTACCGTCCAGGGGTGAACCTCAGC

<210> SEQ ID NO 843

<211> Length : 22

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 843
>HUMCEA_PEA_1_node_33
CGGAGCTGCCCAAGCCCTCCAT

<210> SEQ ID NO 844

<211> Length : 82

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 844
>HUMCEA_PEA_1_node_34
CTCCAGCAACAACTCCAAACCCGTGGAGGACAAGGATGCTGTGGCCTTCACCTGTGAACCTGAGGCTCAGAACACAA
CCTAC

<210> SEQ ID NO 845

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 845
>HUMCEA_PEA_1_node_35
CTGTGGTGGGTAAATGGTCAGAGCCTCCCAGTCAGTCCCAGGCTGCAG

<210> SEQ ID NO 846

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 846
>HUMCEA_PEA_1_node_45
GTAAGTGGCTCCCTGGAGCATCAGCATCATATT

<210> SEQ ID NO 847

<211> Length : 27

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 847
>HUMCEA_PEA_1_node_50
CATCTGGAACTTCTCCTGGTCTCTCAG

<210> SEQ ID NO 848

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 848
>HUMCEA_PEA_1_node_51
CTGGGGCCACTGTCGGCATCATGATTGGAGTGCTGGTTGGGGTTGCTCTGATATAGCAGCCCTGGTGTAGTTTCTTC
ATTTCAGGAAGACTG

<210> SEQ ID NO 849

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 849
>HUMCEA_PEA_1_node_56
ACAGTTGTTTTGCTTCTTCCTTAAAG

<210> SEQ ID NO 850

<211> Length : 101

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 850
>HUMCEA_PEA_1_node_57
CATTTGCAACAGCTACAGTCTAAAATTGCTTCTTTACCAAGGATATTTACAGAAAAGACTCTGACCAGAGATCGAGA
CCATCCTAGCCAACATCGTGAAAC

<210> SEQ ID NO 851

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 851
>HUMCEA_PEA_1_node_58

CCCATCTCTACTAAAAATACAAAAATGAGCTGGG

<210> SEQ ID NO 852

<211> Length : 44

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 852
>HUMCEA_PEA_1_node_60
CTTGGTGGCGCGCACCTGTAGTCCCAGTTACTCGGGAGGCTGAG

<210> SEQ ID NO 853

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 853
>HUMCEA_PEA_1_node_61
GCAG

<210> SEQ ID NO 854

<211> Length : 88

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 854
>HUMCEA_PEA_1_node_62
GAGAATCGCTTGAACCCGGGAGGTGGAGATTGCAGTGAGCCCAGATCGCACCACTGCACTCCAGTCTGGCAACAGAG
CAAGACTCCAT

<210> SEQ ID NO 855

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 855
>HUMCEA_PEA_1_node_64
ATAATATTTTCATAATTTTTTATTTGAAAT

Segment nucleic acid sequences:

<210> SEQ ID NO 856

<211> Length : 266

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 856
>R35137_PEA_1_PEA_1_PEA_1_node_2
TGCCACCTCACCCACTGCCTCTGCCTCCCTGGGGCAGAGCTGTTCCCAGACGGGTGGGGCGGGGCCCAACTGTCCCC
AGCTCCTTCAGCCCTTTCTGTCCCTCCCAGTGAGGCCAGCTGCGGTGAAGAGGGTGCTCTCTTGCCTGGAGTTCCCT
CTGCTACGGCTGCCCCCTCCCAGCCCTGGCCCACTAAGCCAGACCCAGCTGTCGCCATTCCCACTTCTGGTCCTGCC
ACCTCCTGAGCTGCCTTCCCGCCTGGTCTGGGTAG

<210> SEQ ID NO 857

<211> Length : 166

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 857
>R35137_PEA_1_PEA_1_PEA_1_node_3

AGTCATGGCCTCGAGCACAGGTGACCGGAGCCAGGCGGTGAGGCATGGACTGAGGGCGAAGGTGCTGACGCTGGACG
GCATGAACCCGCGTGTGCGGAGAGTGGAGTACGCAGTGCGTGGCCCCATAGTGCAGCGAGCCTTGGAGCTGGAGCAG
GAGCTGCGCCAG

<210> SEQ ID NO 858

<211> Length : 134

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 858
>R35137_PEA_1_PEA_1_PEA_1_node_9
GGGCCTACAGCGTCAGCTCCGGCATCCAGCTGATCCGGGAGGACGTGGCGCGGTACATTGAGAGGCGTGACGGAGGC
ATCCCTGCGGACCCCAACAACGTCTTCCTGTCCACAGGGGCCAGCGATGCCATCGTG

<210> SEQ ID NO 859

<211> Length : 190

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 859
>R35137_PEA_1_PEA_1_PEA_1_node_11
ACGGTGCTGAAGCTGCTGGTGGCCGGCGAGGGCCACACACGCACGGGTGTGCTCATCCCCATCCCCCAGTACCCACT
CTACTCGGCCACGCTGGCAGAGCTGGGCGCAGTGCAGGTGGATTACTACCTGGACGAGGAGCGTGCCTGGGCGCTGG
ACGTGGCCGAGCTTCACCGTGCACTGGGCCAGGCGC

<210> SEQ ID NO 860

<211> Length : 137

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 860

>R35137_PEA_1_PEA_1_PEA_1_node_16
GTGTACCAGGACAACGTGTACGCCGCGGGTTCGCAGTTCCACTCATTCAAGAAGGTGCTCATGGAGATGGGGCCGCC
CTACGCCGGGCAGCAGGAGCTTGCCTCCTTCCACTCCACCTCCAAGGGCTACATGGGCGA

<210> SEQ ID NO 861

<211> Length : 175

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 861
>R35137_PEA_1_PEA_1_PEA_1_node_18
GTGCGGGTTCCGCGGCGGCTATGTGGAGGTGGTGAACATGGACGCTGCAGTGCAGCAGCAGATGCTGAAGCTGATGA
GTGTGCGGCTGTGCCCGCCGGTGCCAGGACAGGCCCTGCTGGACCTGGTGGTCAGCCCGCCCGCGCCCACCGACCCC
TCCTTTGCGCAGTTCCAGGCT

<210> SEQ ID NO 862

<211> Length : 156

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 862
>R35137_PEA_1_PEA_1_PEA_1_node_20
GAGAAGCAGGCAGTGCTGGCAGAGCTGGCGGCCAAGGCCAAGCTCACCGAGCAGGTCTTCAATGAGGCTCCTGGCAT
CAGCTGCAACCCAGTGCAGGGCGCCATGTACTCCTTCCCGCGCGTGCAGCTGCCCCCGCGGGCGGTGGAGCGCGCTC
AG

<210> SEQ ID NO 863

<211> Length : 2,023

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 863

>R35137_PEA_1_PEA_1_PEA_1_node_27

GAGCCCTGGGAGGCTCTGGAGCCCACTGTACTTGCTCTTGATGCCTGGCGGGGTGGGGTGGGGGGGGTGCTGGGCCC
CTGCCTCTCTGCAGGTCCCTAATAAAGCTGTGTGGCAGTCTGACTCCAAAAAGGAAGCGTTGGCAGCTGCGTGGCCC
GCTCCCACCTGCCTACCCTTCTTGCAGGCCTGAGTCCCTTCAGAGAAGGGACCTTCCACGGCCACCACCCACCTCTT
CCTCCTGAAGACCCCGTGCCCACCATAGGCTGGGTCTTCCCTCTGGCCTCTGGTTGTGGGGCAGAGCCCGTCAGATC
ACACAGAAATGGGTTGAGAGGGTCCAGAGTGTGAGGAAAGCGCAGGCCCCACACCCCTTTGTGGAAGCCCCCAAGAA
TCTAGGGAGCCAGGGGCCCAGGTGGCCACCCGAAGAAACACAGCCTTTCCTGAGGAAGGCACAGTGAACTGCCTCCT
TCCTGGCTCCCTTTCCTGTGAGGTCCATGTCTTCCCTGGGGCAGGGGGAAATACTAAACCAGCATGGTGCTGGCTGG
TCAGGGTGACTGACAGCTCAGGAAGGAAGTCTTGGTTCTCTTACCCAAGGAAGCAGGGGTGGGGCCACTGTCTGGGG
GGCCAGAGACCACCTTTGGTGTCATTGTGTGGTGCAGTCCTTCGGCTGGGTGGAGTGGGAGGCAGAGGGAGAGGATA
AGGGAGCGTCCCAGGGGAGGGCTGGGGCTGGAGGAGGCAGGGGCTGGGCTGAGCGGGAGTGGGCAGCGCTGTGTCCT
GGCCTGGGGAGCATGGCTGAGCACCTACTACATGCAGACACTGCTGGGGGGTCACTCAATCTGCACAGATGCTCTTC
TGAAGTAGGCATGATAGTCCCCATTTGATAGACGTGGAAACCTGCAACCCAAAAACCTGCTGAGCTGACAAGAAACC
CCCTCAGAGGCCTCGGCAGCCAGAAAAATGCGTTGGGTCCAGTGCCCTCAAGTCCGCCAAGGACATGGGCTGGCTTT
AGAGACTCACAAACTTGGGAGATAGGACTGGCCAAGGGCACCTGGTTTTTTTCGCTCTGGAGATGGTTCTTAACCAC
AGGCCACACACACTTCACAGCCTCATCTGGCCCTCGGGAGCCCCAGAGGGCACAGCTCTGGGCAGGAGACACAGCAG
GTGGGCCCCTCCCTTGGCAGGGCGGGCTCGAATCAGGCAGGGTGCTCCTAGCCTTGTCACCGGACACCGAAGGGCGT
CACGGGCAGTGGCTGGCGGTGTCCTTCTGAGCTAAGCTGGGTCCTGACCTTTCACACTTCCCTCCTAACTTCCATGG
CTCTGTCACCGCCTTACGGAGGAGCTGAAGCCACAGACACAGCAAGGTTGGGGTCCGCACCGGAAGTATCCAGTGGT
AGACGGCGGAACCCTTAAGAAACGGACGCCTTCATGCGGGCGGCTGGAGAAGCGGGGGCTGGGCACTGCAGCAACCA
CGCTTCGGCTGACACCAGGAAGGAAGCACGCCTGGAGCGGATCCGAAACTACTGAGAGGGGCCAGGGCTGGCCGTGG
GCGCAGGCCGATCCTTACATTCGAGGCCGGCCCAGCTCTGTAGCTTCCCCCTCTGGGCCTCTCACCCGCGCAGGACC
TCGGTGGGAGCGCGCACGTGGCGGGGCGGGGGGCCGCGGGCCCGAGCCCGGACTGGCCACCGGGGGCGCCCGCGAGC
TGACGCTCTGGCCCGTTGCGGTCTCTGTGGCCCGCGCGACCTTCCGGCCCTGGAGCCGGTGGCCGCGGGGCTCCAGC
GACGCCGTGTGGTCCGTGCTCCGCTCTGTGGCTCCAGGGGGCCGAGAAACTGCTGAGAGTCCGGCCCGGCCGGCAGT
GCTGGGCGCGGGCCGAGGGCCCCGGGAGGCAGCGGCCCCGCCCTCTTTACCTGCGGCCTCGCAGAGCATGCTGGGAG
CCGCGGGAGGCAGTGGCCCCGCTCCCCTCACCTGCGGTATCGCAGAGCATGGTGGGAGCCCCGGGAGGCAGTGGCCC
CGCCCCCTTCCCTGCGGCCGC

<210> SEQ ID NO 864

<211> Length : 90

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 864

>R35137_PEA_1_PEA_1_PEA_1_node_5
GGTGTGAAGAAGCCTTTCACCGAGGTCATCCGTGCCAACATCGGGGACGCACAGGCTATGGGGCAGAGGCCCATCAC
CTTCCTGCGCCAG

<210> SEQ ID NO 865

<211> Length : 109

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 865
>R35137_PEA_1_PEA_1_PEA_1_node_7
GTCTTGGCCCTCTGTGTTAACCCTGATCTTCTGAGCAGCCCCAACTTCCCTGACGATGCCAAGAAAAGGGCGGAGCG
CATCTTGCAGGCGTGTGGGGGCCACAGTCTGG

<210> SEQ ID NO 866

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 866
>R35137_PEA_1_PEA_1_PEA_1_node_12
GTGACCACTGCCGCCCTCGTGCGCTCTGTGTCATCAACCCTGGCAACCCCACCG

<210> SEQ ID NO 867

<211> Length : 80

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 867
>R35137_PEA_1_PEA_1_PEA_1_node_14

GGCAGGTGCAGACCCGCGAGTGCATCGAGGCCGTGATCCGCTTCGCCTTCGAAGAGCGGCTCTTTCTGCTGGCGGAC
GAG

<210> SEQ ID NO 868

<211> Length : 67

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 868
>R35137_PEA_1_PEA_1_PEA_1_node_15
GTGCGCGGCGCGGGGGAGCGGGAAGCCGGGCAACAGTCCGCCCCCGTGACGCCTTGCGCCCTTCCAG

<210> SEQ ID NO 869

<211> Length : 100

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 869
>R35137_PEA_1_PEA_1_PEA_1_node_17
GTGCGTGCGTACGAGGCGGGTGGGGGCTCGCGGGCCATGGCCAGGCCCTCCTCGCCCGATGGGCCACCCCCTCCTCC
GCACCTGACCTGGCCGTGCGCAG

<210> SEQ ID NO 870

<211> Length : 74

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 870
>R35137_PEA_1_PEA_1_PEA_1_node_21
GTCAGGCGGGGGCGGGGCCTGCGGGGTGGGCAGGGGGGGCCGGGCATCCCTCTCTGACGGCTCTCCGTCCACAG

<210> SEQ ID NO 871

<211> Length : 73

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 871
>R35137_PEA_1_PEA_1_PEA_1_node_22
GAGCTGGGCCTGGCCCCCGATATGTTCTTCTGCCTGCGCCTCCTGGAGGAGACCGGCATCTGCGTGGTGCCAG

<210> SEQ ID NO 872

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 872
>R35137_PEA_1_PEA_1_PEA_1_node_23
GGAGCGGCTTTGGGCAGCGGGAAGGCACCTACCACTTCCG

<210> SEQ ID NO 873

<211> Length : 66

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 873
>R35137_PEA_1_PEA_1_PEA_1_node_24
GTGAGGCCTGGCCCTCACTCCCTGTCCCGCCACCCTGGCCCTTCACTCACTGTCAACTCCTTTCAG

<210> SEQ ID NO 874

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 874
>R35137_PEA_1_PEA_1_PEA_1_node_25
GATGACCATTCTGCCCCCCTTGGAGAAACTGCGGCTGCTGCTGGAGAAGCTGAGCAGGTTCCATGCCAAGTTCACCC
TCGA

<210> SEQ ID NO 875

<211> Length : 57

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 875
>R35137_PEA_1_PEA_1_PEA_1_node_26
GTACTCCTGAGCACCCCAGCTGGGGCCAGGCTGGGTCGCCCTGGACTGTGTGCTCAG

Segment nucleic acid sequences:

<210> SEQ ID NO 876

<211> Length : 582

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 876
>Z25299_PEA_2_node_20
GTAAGCAGGGGATGAGGGCACACTGAGCTCCCTCCAGCCCTCTCAGCCTCAACCCTCTGGAGGCCCAGGCATATGGG
CAGGGGGACTCCTGAACCCTACTCCAAGCACAGCCTCTGTCTGACTCCCTTGTCCTTCAAGAGAACTGTTCTCCAGG
TCTCAGGGCCAGGATTTCCATAGGATCGCCTGTGGCTTTGATTCTATTCTAGTGTCTCTGGGTGGGGGTCCTGGGCA
AGTGTCTTTCTGAGTCTCAGTTTCTTTATCGGTAAAATGTACATAATGAGATTGAAAGTGCTCTGCAAAGCACTATG
TGCACTAAGAATTTATTATTCAGGTTGTTTCCATCATGTTTTCTGAGGTGAAATCACAAAGGATCAGTGGAGTTTGA
GGATTATCTAGTTCAATGCTTTGAGTTTAGAGTTTTACGGTGAAAATGAGACTTGTCTCCTGACACTAAGTCTCTCT

CAACTATAGCGCTATCTTGCTATTTTCTCTATCTCAGAAGGATCCTTGGGCAGGAGGAAGGATGTGGATATGATTTG
GCTGGTTTCTATGCTGAAGCTCTTATCTGATTTTCTCTCACAG

<210> SEQ ID NO 877

<211> Length : 193

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 877
>Z25299_PEA_2_node_21
CTTGATTCCTGCCATATGGAGGAGGCTCTGGAGTCCTGCTCTGTGTGGTCCAGGTCCTTTCCACCCTGAGACTTGGC
TCCACCACTGATATCCTCCTTTGGGGAAAGGCTTGGCACACAGCAGGCTTTCAAGAAGTGCCAGTTGATCAATGAAT
AAATAAACGAGCCTATTTCTCTTTGCAAAACCTGCTTCT

<210> SEQ ID NO 878

<211> Length : 190

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 878
>Z25299_PEA_2_node_23
GTGAAAAGAGACATCACAAGCAATTGAGGGACCAGGAAGTGGATCCTCTAGAGATGAGGAGGCATTCTGCTGGATGA
CTTTTAAAAATGTTTTCTCCAGAGTCATCTCTCTCATTAACAATGTTTTTTGTCTTAGAAATTTCTTGTTGATTTTT
AAACTTACATGATTTCTTGTTTTGGTATGAATACAG

<210> SEQ ID NO 879

<211> Length : 179

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 879

>Z25299_PEA_2_node_24
GCTGCTTCAGTCCTTCAATAAGCCCATCACACTTTTTCACCATGTCATCTATCAGCACTTTTTCTGCAGTGTTACGA
ACATCAGCTTCATCACTGTCAGCCTGCGTTTTGCCTGCAACCCATCAAATGAGGTCAGGAGAGGAGTTTTCCACTTT
TGGCTTCATGTTGGTGCTCAAAACT

<210> SEQ ID NO 880

<211> Length : 208

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 880
>Z25299_PEA_2_node_8
GTTTCCTGCTTATGCAATAGTAGCTGGGAGAGGCCGAAAGAATTCTGGTGGGGCCACACCCACTGGTGAAAGAATAA
ATAGTGAGGTTTGGCATTGGCCATCAGAGTCACTCCTGCCTTCACCATGAAGTCCAGCGGCCTCTTCCCCTTCCTGG
TGCTGCTTGCCCTGGGAACTCTGGCACCTTGGGCTGTGGAAGGCTCTGGAAAGT

<210> SEQ ID NO 881

<211> Length : 37

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 881
>Z25299_PEA_2_node_12
CCTTCAAAGCTGGAGTCTGTCCTCCTAAGAAATCTGC

<210> SEQ ID NO 882

<211> Length : 112

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 882

>Z25299_PEA_2_node_13
CCAGTGCCTTAGATACAAGAAACCTGAGTGCCAGAGTGACTGGCAGTGTCCAGGGAAGAAGAGATGTTGTCCTGACA
CTTGTGGCATCAAATGCCTGGATCCTGTTGACACC

<210> SEQ ID NO 883

<211> Length : 10

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 883
>Z25299_PEA_2_node_14
CCAAACCCAA

<210> SEQ ID NO 884

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 884
>Z25299_PEA_2_node_17
CAAG

<210> SEQ ID NO 885

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 885
>Z25299_PEA_2_node_18
GAGGAAGCCTGGGAAGTGCCCAGTGACTTATGGCCAATGTTTGATGCTTAACCCCC

<210> SEQ ID NO 886

<211> Length : 90

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 886
>Z25299_PEA_2_node_19
CCAATTTCTGTGAGATGGATGGCCAGTGCAAGCGTGACTTGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTT
TCCCCTGTGAAAG

Segment nucleic acid sequences:

<210> SEQ ID NO 887

<211> Length : 131

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 887
>HSSTROL3_node_6
CAAGCCCAGCAGCCCCGGGGCGGATGGCTCCGGCCGCCTGGCTCCGCAGCGCGGCCGCGCGCGCCCTCCTGCCCCCG
ATGCTGCTGCTGCTGCTCCAGCCGCCGCCGCTGCTGGCCCGGGCTCTGCCGCCG

<210> SEQ ID NO 888

<211> Length : 182

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 888
>HSSTROL3_node_10

GACGTCCACCACCTCCATGCCGAGAGGAGGGGGCCACAGCCCTGGCATGCAGCCCTGCCCAGTAGCCCGGCACCTGC
CCCTGCCACGCAGGAAGCCCCCCGGCCTGCCAGCAGCCTCAGGCCTCCCCGCTGTGGCGTGCCCGACCCATCTGATG
GGCTGAGTGCCCGCAACCGACAGAAGAG

<210> SEQ ID NO 889

<211> Length : 144

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 889
>HSSTROL3_node_13
GATCCTTCGGTTCCCATGGCAGTTGGTGCAGGAGCAGGTGCGGCAGACGATGGCAGAGGCCCTAAAGGTATGGAGCG
ATGTGACGCCACTCACCTTTACTGAGGTGCACGAGGGCCGTGCTGACATCATGATCGACTTCGCCAG

<210> SEQ ID NO 890

<211> Length : 134

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 890
>HSSTROL3_node_15
GTACTGGCATGGGGACGACCTGCCGTTTGATGGGCCTGGGGGGCATCCTGGCCCATGCCTTCTTCCCCAAGACTCACC
GAGAAGGGGATGTCCACTTCGACTATGATGAGACCTGGACTATCGGGGATGACCAGG

<210> SEQ ID NO 891

<211> Length : 183

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 891
>HSSTROL3_node_19

ACAACAGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGACTGCAG
GGGCGTTCAACACCTATATGGCCAGCCCTGGCCCACTGTCACCTCCAGGACCCCAGCCCTGGGCCCCCAGGCTGGGA
TAGACACCAATGAGATTGCACCGCTGGAG

<210> SEQ ID NO 892

<211> Length : 217

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 892
>HSSTROL3_node_21
CCAGACGCCCCGCCAGATGCCTGTGAGGCCTCCTTTGACGCGGTCTCCACCATCCGAGGCGAGCTCTTTTTCTTCAA
AGCGGGCTTTGTGTGGCGCCTCCGTGGGGGCCAGCTGCAGCCCGGCTACCCAGCATTGGCCTCTCGCCACTGGCAGG
GACTGCCCAGCCCTGTGGACGCTGCCTTCGAGGATGCCCAGGGCCACATTTGGTTCTTCCAAG

<210> SEQ ID NO 893

<211> Length : 138

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 893
>HSSTROL3_node_24
AGCTGGGATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCCAGGGACTTCACAAATGAAGG
CACAGCATGGGAAACCTGCGTGGGTTCCAGGGCAGTCCAGCCTGCAGGGGCCCAGGGAGTG

<210> SEQ ID NO 894

<211> Length : 300

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 894

>HSSTROL3_node_25
GTCAGTAGGCATTTGTCACAGCCAAATGCCAGTGGAAGGAGCAGCCGCCCAGGCAGCCCTCTACTGATGAGAGTAAC
CTCACCCGTGCACTAGTTTACAGAGCATTCACTGCCCCAGCTTATCCCAGGCCTCCCGCTTCCCTCTGCGGGTGGGG
TGCTGAGCAGGCATTATTGGCCTGCATGTTTTACTGATGAGGAAACTGAGGCTGGGAGAGTCTGTGGTAGGGGTCAA
GCAGGTCCACAGTGGCGGGGCATGGCAGTGGTGGCTGGGCAGGTCCTTGCAGCCTTCCCTCTCCGGCAG

<210> SEQ ID NO 895

<211> Length : 142

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 895
>HSSTROL3_node_26
GTGCTCAGTACTGGGTGTACGACGGTGAAAAGCCAGTCCTGGGCCCCGCACCCCTCACCGAGCTGGGCCTGGTGAGG
TTCCCGGTCCATGCTGCCTTGGTCTGGGGTCCCGAGAAGAACAAGATCTACTTCTTCCGAGGCAG

<210> SEQ ID NO 896

<211> Length : 927

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 896
>HSSTROL3_node_28
GTGCGTTGGGGGTGAGGCAGCTGGTGGGAGGTGGGCACAGCAGCCGCTTCTCCCACCTGGTGGTGGCTGGGCTCCCA
CATGCCTGCCACAGGAAGTCTGGCTCTTCATCACAGGTCCTTTGTCCAGAGCCATCTGCCCTCCTCTCGGTGGCCGG
CTAGTGCTACATTCCATATTGCAGATGAGGAAACTGAGGGTCAGAGAAGTGCAAGGTCTTACCCTGGTTTTTCAGCC
ACAGCCAGTAGAACAATAAACTGCTGTACACTGAGGGCCAACAATGCTCTAAGCTCCTTACTGGTCTCATCCAGTTC
TCAGAACAGCCCTCTGATGTGACACCTGTTGTGAACCCAGTTTCCAGAGGAGCAAACAGAGGCTCAGGCAATGAGGC
CCCTAACCTGGACTACCCTGGTGGTCCCTGCTCCTAACCACTGACCCACCCAGCCTCCCACAACCACAGGGGGCTAG
AGCCAGTCCAGTGCTCCCTCCCCTGCTAGGCTCCTCTTCTGTGCTCTTTCTCCCACATCAGGACCCACTGGGAGAGC
TATCCTAGGGTAGCCTCCAGCTCCAGGACTCCAGGGTGCCCGTCAATAGCCTGGCTAATTTAATAGATGCAGGAGAG
AGTGATGTGGAGGGTGGTGGGGGCAACGGGACTTGCTTTCCTGAGAGGTGGGACTCAGGCCTCTGAGGCTCTGGGTA
CCTGTCAGGCTGGGTATTAGCCCAGCCCAGATTCCGGGGCAGGCAGAAGGGCTCCCTAGAGGGAAGAGAGGTTCTGA
AAGGCCGGCCCTGGATCCTGCAGGACTCGAGGAACTCAGCAGTGGCCAAGGGCTTCCCACTCAGCCCTCCCTTAGTG

CCCATCCCTGGGCACAGCCTGACAGGCAGGAGTAGGGCCCAGTGTCCACTCGCCCAGGCTTGACCACCTTCTCTTCT
CAG

<210> SEQ ID NO 897

<211> Length : 911

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 897
>HSSTROL3_node_29
GCTATGCCTACTTCCTGCGCGGCCGCCTCTACTGGAAGTTTGACCCTGTGAAGGTGAAGGCTCTGGAAGGCTTCCCC
CGTCTCGTGGGTCCTGACTTCTTTGGCTGTGCCGAGCCTGCCAACACTTTCCTCTGACCATGGCTTGGATGCCCTCA
GGGGTGCTGACCCCTGCCAGGCCACGAATATCAGGCTAGAGACCCATGGCCATCTTTGTGGCTGTGGGCACCAGGCA
TGGGACTGAGCCCATGTCTCCTCAGGGGGATGGGGTGGGGTACAACCACCATGACAACTGCCGGGAGGGCCACGCAG
GTCGTGGTCACCTGCCAGCGACTGTCTCAGACTGGGCAGGGAGGCTTTGGCATGACTTAAGAGGAAGGGCAGTCTTG
GGCCCGCTATGCAGGTCCTGGCAAACCTGGCTGCCCTGTCTCCATCCCTGTCCCTCAGGGTAGCACCATGGCAGGAC
TGGGGGAACTGGAGTGTCCTTGCTGTATCCCTGTTGTGAGGTTCCTTCCAGGGGCTGGCACTGAAGCAAGGGTGCTG
GGGCCCCATGGCCTTCAGCCCTGGCTGAGCAACTGGGCTGTAGGGCAGGGCCACTTCCTGAGGTCAGGTCTTGGTAG
GTGCCTGCATCTGTCTGCCTTCTGGCTGACAATCCTGGAAATCTGTTCTCCAGAATCCAGGCCAAAAAGTTCACAGT
CAAATGGGGAGGGGTATTCTTCATGCAGGAGACCCCAGGCCCTGGAGGCTGCAACATACCTCAATCCTGTCCCAGGC
CGGATCCTCCTGAAGCCCTTTTCGCAGCACTGCTATCCTCCAAAGCCATTGTAAATGTGTGTACAGTGTGTATAAAC
CTTCTTCTTCTTTTTTTTTTTTAAACTGAGGATTGTCATTAAACACAGTTGTTTTCTACCTGCC

<210> SEQ ID NO 898

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 898
>HSSTROL3_node_11
GTTCGTGCTTTCTGGCGGGCGCTGGGAGAAGACGGACCTCACCTACAG

<210> SEQ ID NO 899

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 899
>HSSTROL3_node_17
GCACAGACCTGCTGCAGGTGGCAGCCCATGAATTTGGCCAC

<210> SEQ ID NO 900

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 900
>HSSTROL3_node_18
GTGCTGGGGCTGCAGCAC

<210> SEQ ID NO 901

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 901
>HSSTROL3_node_20
GTGAGGCCCTGCCTGCCAGTCCCCCTACTCCTCTGCTGGCCACTGTGACTGCAGCATATGCCCTCAGCATGTGTCCC
TCTCTCCCACCCCAG

<210> SEQ ID NO 902

<211> Length : 116

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 902
>HSSTROL3_node_27
GGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAGGGCCACTGACTGGAGAGGGG
TGCCCTCTGAGATCGACGCTGCCTTCCAGGATGCTGATG

Segment nucleic acid sequences:

<210> SEQ ID NO 903

<211> Length : 359

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 903
>HUMTREFAC_PEA_2_node_0
CGCTCCCCAGTAGAGGACCCGGAACCAGAACTGGAATCCGCCCTTACCGCTTGCTGCCAAAACAGTGGGGGCTGAAC
TGACCTCTCCCCTTTGGGAGAGAAAAACTGTCTGGGAGCTTGACAAAGGCATGCAGGAGAGAACAGGAGCAGCCACA
GCCAGGAGGGAGAGCCTTCCCCAAGCAAACAATCCAGAGCAGCTGTGCAAACAACGGTGCATAAATGAGGCCTCCTG
GACCATGAAGCGAGTCCTGAGCTGCGTCCCGGAGCCCACGGTGGTCATGGCTGCCAGAGCGCTCTGCATGCTGGGGC
TGGTCCTGGCCTTGCTGTCCTCCAGCTCTGCTGAGGAGTACGTGGGCCTGT

<210> SEQ ID NO 904

<211> Length : 586

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 904
>HUMTREFAC_PEA_2_node_9

CTCCAGCTGCCCCCGGCCGGGGGATGCGAGGCTCGGAGCACCCTTGCCCGGCTGTGATTGCTGCCAGGCACTGTTCA
TCTCAGCTTTTCTGTCCCTTTGCTCCCGGCAAGCGCTTCTGCTGAAAGTTCATATCTGGAGCCTGATGTCTTAACGA
ATAAAGGTCCCATGCTCCACCCGAGGACAGTTCTTCGTGCCTGAGACTTTCTGAGGTTGTGCTTTATTTCTGCTGCG
TCGTGGGAGAGGGCGGGAGGGTGTCAGGGGAGAGTCTGCCCAGGCCTCAAGGGCAGGAAAAGACTCCCTAAGGAGCT
GCAGTGCATGCAAGGATATTTTGAATCCAGACTGGCACCCACGTCACAGGAAAGCCTAGGAACACTGTAAGTGCCGC
TTCCTCGGGAAAGCAGAAAAAATACATTTCAGGTAGAAGTTTTCAAAAATCACAAGTCTTTCTTGGTGAAGACAGCA
AGCCAATAAAACTGTCTTCCAAAGTGGTCCTTTATTTCACAACCACTCTCGCTACTGTTCAATACTTGTACTATTCC
TGGGTTTTGTTTCTTTGTACAGTAAACATTATGAACAAACAGGCAAA

<210> SEQ ID NO 905

<211> Length : 111

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 905
>HUMTREFAC_PEA_2_node_2
GGAAAGTGCATCTTCCTAAGGGCGAGGGTTTCAGCAGTGGTTGAACTCGGCGGGGTGGGGCGGAGCGGGAGGATGCA
AACTTGCAAAGTGAAGCAAACACACTCACCGCAG

<210> SEQ ID NO 906

<211> Length : 44

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 906
>HUMTREFAC_PEA_2_node_3
CCCAGCAAGGGCTCTGGCAGCTGACAGGGCTTTGTCTGGGACAG

<210> SEQ ID NO 907

<211> Length : 97

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 907
>HUMTREFAC_PEA_2_node_4
CTGCAAACCAGTGTGCCGTGCCAGCCAAGGACAGGGTGGACTGCGGCTACCCCCCATGTCACCCCCAAGGAGTGCAAC
AACCGGGGCTGCTGCTTTGA

<210> SEQ ID NO 908

<211> Length : 50

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 908
>HUMTREFAC_PEA_2_node_5
CTCCAGGATCCCTGGAGTGCCTTGGTGTTTCAAGCCCCTGCAGGAAGCAG

<210> SEQ ID NO 909

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 909
>HUMTREFAC_PEA_2_node_8
AATGCACCTTCTGAGGCAC

Segment nucleic acid sequences:

<210> SEQ ID NO 910

<211> Length : 1,133

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 910

>HSS100PCB_node_3

TGAGACAAGATGTCACTCTGTCACCCAGGCTGGAGTGCAGTGGCAGGATCACGGCTCACTGCAGCCTCGACCTCCCT
GGGCTCAGGTGATCCTCCCACCTCAGCCTACCGAGTAGCTGGGACTACAGGTGCATGTCACCATACCCGGTTAATTT
TTGTATTTTTTTTAGAGACAAGGTCTCACCATGTTGCCCAGGCTGGTCTCAAACTCCTGTGCTCAGGCAATGCGTCA
GCCTCGACATCTCAAAGTGCTGTGATTACAGGCGTGAGCCCCGACACCTGGCCTAGTTCTATTTTCTAAATGTGAAT
TCTGTAAAGATATCTTTTAAAAATAAAGTTCTGTTTTTGGTAGAAAATGTAAAAATAGATAAATATGGAGGGAAGAA
ATCCCCCCTGGAATACAGACGCTTCCTCTCCCTTCCAGCCTTTTCCCCATATGAACATTGCTGTGAGTGAGACTTAC
ATGCAATGTAATTTCTTTTTGAGCTTAACATTACAACATAAATTCTCAAACTCTGATGTTCATTAAACACCCCAGCC
CCATCCTGGGAACTTGGGCTTGGGGCTCGGGGTGTTCTGATAATGATCAAAGTATGAGAATTGAACCCATGAGGACT
TTGATCCAAGATACTGGGGTGTGGGGAGGGGCAGGCACAGGTGTCCTGGGAACACACTTTGAGAAGCAATGGCAAAG
CTGGGGGTCCAGCTAATGTGTTACATTAGAATCACCTCGGGGAGGCCCTGGGTGCCCTTCTCAGCCCTCCCTCCGGA
GGCTGCTGAAGCCCAGCAAAGCCGGAGTCAGAGAACAATGTCCGCCTGAGGGCAGGGCTGGGCTGGGCTGGCCTTCT
GGCCCTATCTGCTCCGTGCCCAACCCAGCGCCCCGCACAGTCGGAGCTTTGTAAATACGAGGTGACTGTCTGCCTAC
AAACTTTGTAAACATCACTTGAAATGGCCGCAGGGCATTGCGACATGGCCATACCACTATTTGTTTGCTATTGAATT
TGTACTTCCCTGCCTTACTTTTGCTATTGCAAACCATGCTGTCACTAAGGTCTTCATGCACACAGTTGTGTCTTGGT
CAGATGATATGTTTCTACCAATTTTAATTGTGTTTCTTTCCACCTGGACACACAG

<210> SEQ ID NO 911

<211> Length : 790

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 911

>HSS100PCB_node_4

CTCTCTGGCCCAGGGCTGGGTCATCAGCACACCCTGCTGCTGCTGTTCAGATCTGCATCCTGGTCCCGCTTGGTCCC
ACAGTGAGAACGCTTTGCTATCACATGGGCAGGCTCTGAGAGCCCTGCCGGCCTGGCCTTCTCAAAGAAGACCTGAG
AGCTTGGGACCCAAGCAGAGAGGAAGAACAGGGCTCAGGGTGCTTGCTCCATGCTCGCTCCACACCTGGGGCTCAAC
CCTGGCTTTCCCCGGCTCCCTGTGTGACTTCAGGGCAGGTCCCTTGGGCCCTCTGGGCCTTATCATCTTCATCTGTA
ACAGGGCGATGCCTCTGCCGTGTCTGGTGGTGTTGAGGAGTTCCTGTTTGTGTAAGCAGCTAGTTCAGTGCCAGCAC
GAGATGGGAGGCCCATGAAGTTAGCAGTGCACAAAAAATAGAGCAAAGACTGGATGCATTTCCTGAGAACAACCATC
ACTGTAAAGCACTTTACAAATCCAAAGACAACCCCCGGCAAAAACTCAAAATGAAACTCCCTCTCGCAGAGCACAAT
TCCAATTCGCTCTAAAAACATTACAAGTTAGTTCATGTCATGCCAGATAGCTGAAGGCAGCTCACAAGTTCTTAAGG

CCAGGAATGCCATGTGTCTGCTATGCACAGCTGGCCCTGGCCCTGAGCCTGAATGACAGCACAAAGGTGACGCAGAT
GTGGGTGCCCTGCTCCTGCCCAGCAGCAGTGCTTGGTGGAGGCTGAGGCCCTGCACAGGCACCCTCACTGCTGACCT
TGAGCCTCTCTCTCCTCTAG

<210> SEQ ID NO 912

<211> Length : 643

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 912
>HSS100PCB_node_5
AGTGGAAAAGACAAGGATGCCGTGGATAAATTGCTCAAGGACCTGGACGCCAATGGAGATGCCCAGGTGGACTTCAG
TGAGTTCATCGTGTTCGTGGCTGCAATCACGTCTGCCTGTCACAAGTACTTTGAGAAGGCAGGACTCAAATGATGCC
CTGGAGATGTCACAGATTCCTGGCAGAGCCATGGTCCCAGGCTTCCCAAAAGTGTTTGTTGGCAATTATTCCCCTAG
GCTGAGCCTGCTCATGTACCTCTGATTAATAAATGCTTATGAAATGATCCTGGTCTCAGAGAAACTGGTTATTTCCT
TAAAACCTCAGTGTCCTCTCCCCACAAAACAGAATTCACTAATTAAACTTGTCATTTCATTACTCATTGCCTGCCTG
CAGTGGTGAGACAGATGAAGAAGACAAGGCAGTCGATAAAAATCCATTTGTTGGCCATCTTCTGAGTGCCCACTGAG
ACACAGTCCCACCTTTGGCCGTGTGAATGCCACCATGTCCTGGACCCAAGGGAGGCTCGGGTACACACATCCATGTA
TTTATTCATCTACTTCGTAATTCATTTAGTCGGTCACGTAGCCACTCATTGAACCTGGGCCCCTCCATCTCTCCAAA
CAGCATGTTTTCTTCAAGTAACATACT

Segment nucleic acid sequences:

<210> SEQ ID NO 913

<211> Length : 1,298

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 913
>R20779_node_0
CTGTCCCCGGCCCCGGCGCGGGGGAGACGTGAGCGTGCACACGTACACACACAGCAGGGGAAGAGGCGCTCCAAGCG
GCGCCCAACTTTCTCCTTCCCTCCACGGGCCGGGTGAGAAAGTAGCCGGGGGCTATCCCGACCCGGCGGTTCTTGGG
GAGGGGGCCGAACAAGAAAAGGGAGGAGATGGAGATAACTTCCCCGGATTTAGCTTTTTTGTCTTTGTTTTTGTTCT
CACCACTTCCATCGGATGACTGGAGAGTAAAAGGGAACCCGGAGCGGGGTGGCGAGCAGCGCTTTGAGAAAATGCAG

GAGTGTGTTTGGAGACGCGTAAAGTTGCCTTTCAAGCTCTGGCCTCCGGGCACGCGATGCTCCGCGGCGGGCTGACT
CAGGGCTGCCTTGGGCCTCCCTGCCACCCTCCTGGAAATGATGCAAGTCCTGACTGTCACCTGGATCCCTGCAGCCC
AGCCTGGAATGCGTCTGGATTAGGGGAAAGACGAGAAACGACACTCCAGGTGTTGCACGGCCCACCAAAGCGGGAAG
ATAGGGCAGTTGCTCAGACCAAATACTGTATCTAGTGCTTCTGCTCCTATCTTCAATCGTGGGGTTCTTTTTAATGC
AAAGTGTCACAAGGCCAGGAATTCCCATGTGTGCTCAGTTGGCCCACAGCATCATTGTGCCTAGGAAACTGCTTCAA
TTTATCAAGTCCTCTGGGCTGGGAATCTCACTGAATTCCAAACGGCGGAAAGAGGAAACTTTCCCAACCCGATGTGG
GTGTGACGCGAGCCAGGGGCCCCAGGGACACTGTCCCAGAGCACACCGTCCCCCTTTAACAGCAACTGGAGCTTGGA
TTCGCTCTTATATTGTACAGTCCTTTCGACCATTGCCCTGGAGCACCCGCACACGCGCACGCATCTCCGGCCGCGCT
CACACACACTCATACACACGCACGCAAACGCGTGGCCGCCGCCAGGTCGGCAACTTTGTCCGGCGCTCCCAGCGGCG
CTCGGCTTCCTCCTGTAGTAGTTGAGCGCAGGCCCCGCCTCCCGGCCGTGTTGTCAAAAGGGCCGGGGTCTCGGATT
GGTCCAGCCGCCGGGACAACACCTGCTCGACTCCTTCATTCAAGTGACACCAGAGCTTCCAGGGATATTTGAGGCAC
CATCCCTGCCATTGCCGGGCACTCGCGGCGCTGCTAACGGCCTGGTCACATGCTCTCCGGAGAGCTACGGGAGGGCG
CTGGGTAACCTCTATCCGAGCCGCGGCCGCGAGGAGGAGGGAAAAGGCGAGCAAAAAGGAAGAGTG

<210> SEQ ID NO 914

<211> Length : 170

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 914
>R20779_node_2
GGAGGAAGAGGGGAGCACAAAGGATCCAGGTCTCCCGACGGGAGGTTAATACCAAGAACCATGTGTGCCGAGCGGCT
GGGCCAGTTCATGACCCTGGCTTTGGTGTTGGCCACCTTTGACCCGGCGCGGGGGACCGACGCCACCAACCCACCCG
AGGGTCCCCAAGACAG

<210> SEQ ID NO 915

<211> Length : 143

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 915
>R20779_node_7
CGGAGATCCAGCACTGTTTGGTCAACGCTGGCGATGTGGGGTGTGGCGTGTTTGAATGTTTCGAGAACAACTCTTGT
GAGATTCGGGGCTTACATGGGATTTGCATGACTTTTCTGCACAACGCTGGAAAATTTGATGCCCAG

<210> SEQ ID NO 916

<211> Length : 148

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 916

>R20779_node_9

GGCAAGTCATTCATCAAAGACGCCTTGAAATGTAAGGCCCACGCTCTGCGGCACAGGTTCGGCTGCATAAGCCGGAA

GTGCCCGGCCATCAGGGAAATGGTGTCCCAGTTGCAGCGGGAATGCTACCTCAAGCACGACCTGTGCGCGG

<210> SEQ ID NO 917

<211> Length : 168

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 917

>R20779_node_18

CTGTGGGGAGGAGGTGAAGGAGGCCATCACCCACAGCGTGCAGGTTCAGTGTGAGCAGAACTGGGGAAGCCTGTGCT

CCATCTTGAGCTTCTGCACCTCGGCCATCCAGAAGCCTCCCACGGCGCCCCCCGAGCGCCAGCCCCAGGTGGACAGA

ACCAAGCTCTCCAG

<210> SEQ ID NO 918

<211> Length : 578

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 918

>R20779_node_21

CAGTAGGGAGACTGGCCGAGGTGCCAAGGGTGAGCGAGGTAGCAAGAGCCACCCAAACGCCCATGCCCGAGGCAGAG

TCGGGGGCCTTGGGGCTCAGGGACCTTCCGGAAGCAGCGAGTGGGAAGACGAACAGTCTGAGTATTCTGATATCCGG

AGGTGAAATGAAAGGCCTGGCCACGAAATCTTTCCTCCACGCCGTCCATTTTCTTATCTATGGACATTCCAAAACAT
TTACCATTAGAGAGGGGGGATGTCACACGCAGGATTCTGTGGGGACTGTGGACTTCATCGAGGTGTGTGTTCGCGGA
ACGGACAGGTGAGATGGAGACCCCTGGGGCCGTGGGGTCTCAGGGGTGCCTGGTGAATTCTGCACTTACACGTACTC
AAGGGAGCGCGCCCGCGTTATCCTCGTACCTTTGTCTTCTTTCCATCTGTGGAGTCAGTGGGTGTCGGCCGCTCTGT
TGTGGGGGAGGTGAACCAGGGAGGGGCAGGGCAAGGCAGGGCCCCCAGAGCTGGGCCACACAGTGGGTGCTGGGCCT
CGCCCCGAAGCTTCTGGTGCAGCAGCCTCTGGTGCTGTC

<210> SEQ ID NO 919

<211> Length : 691

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 919

>R20779_node_24

GGAGTGTCATTTCTATGTGTAATTTCTGAGCCATTGTACTGTCTGGGCTGGGGGGGACACTGTCCAAGGGAGTGGCC
CCTATGAGTTTATATTTTAACCACTGCTTCAAATCTCGATTTCACTTTTTTTATTTATCCAGTTATATCTACATATC
TGTCATCTAAATAAATGGCTTTCAAACAAAGCAACTGGGTCATTAAAACCAGCTCAAAGGGGGTTTAAAAAAAAAAA
ACCAGCCCATCCTTTGAGGCTGATTTTTCTTTTTTTTAAGTTCTATTTTAAAAGCTATCAAACAGCGACATAGCCAT
ACATCTGACTGCCTGACATGGACTCCTGCCCACTTGGGGGAAACCTTATACCCAGAGGAAAATACACACCTGGGGAG
TACATTTGACAAATTTCCCTTAGGATTTCGTTATCTCACCTTGACCCTCAGCCAAGATTGGTAAAGCTGCGTCCTGG
CGATTCCAGGAGACCCAGCTGGAAACCTGGCTTCTCCATGTGAGGGGATGGGAAAGGAAAGAAGAGAATGAAGACTA
CTTAGTAATTCCCATCAGGAAATGCTGACCTTTTACATAAAATCAAGGAGACTGCTGAAAATCTCTAAGGGACAGGA
TTTTCCAGATCCTAATTGGAAATTTAGCAATAAGGAGAGGAGTCCAAGGGGACAAATAAAGGCAGAGAGAAGAGA

<210> SEQ ID NO 920

<211> Length : 131

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 920

>R20779_node_27

CTAAAAATACGAGGAAAGGAGAGTGAGGATTTTCATTAAAAGTCTCAGCAGTGGGTTTCTTGGGTTATTTAAAACAT
CACCTAAATAGGCCTTTTCTTCCTAATTGGCCATCAAATTAAAGCCTATCCTTT

<210> SEQ ID NO 921

<211> Length : 247

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 921
>R20779_node_28
CTCAAGCAGGAGCTGGTATTGTAGGGAGTGGCCGGGTATTCTGGGCTGGGCTCTTCTGGAGTAGGGGGTCAGCAAAC
ATTGTCTGCAAAGGGCCAGATACTGAATCCAGTACTTTCAGTTTGGCGAGCCGTGAGGTCTCTGTCGAAACTACTCA
ACTCTGCCGTCCTAGCACAAAAGCAGCCATAGACAACACACAAACGAGAGGGCTTGGCTCCCTTCCAGGAAGATTTA
TTTAACAGGCTCCCAG

<210> SEQ ID NO 922

<211> Length : 126

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence :922
>R20779_node_30
GACACAGTTTACTGATCTCTGTTCTAGTGAGTGGGTCAAAAAGCATATGCATCCTTATCCGTCAACTCATCAGCTCT
TCCTCAAGGCAACCTGAGGCCAGACACCAAGAAACCAAGCGTATCTGCT

<210> SEQ ID NO 923

<211> Length : 231

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 923
>R20779_node_31
CTAAAATGACTTGTTCCTGGGGAATGCCTTCAACCAAAACACAGCTAGTATTTCTATGCCCCAAATCCAATCCCAGT
CTTTCATGATCCATGCCGGCGGTTGGGTGGGGAGGGGAATCATTGGTTGGGGGAAGGGAGGAAACCCCACCTCCAGC
CCCCGCCACCGGGCTCCCTGGGCACCCAGCAAGATCTGGGGCTGCAGAGAACAGAAGAGCTGGTGCACTTAATCCAG

<210> SEQ ID NO 924

<211> Length : 1,079

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 924

>R20779_node_32

CTCTGCCCTTGGGGGGAGGAGGACCTGTGTGTCAGGCTCTGCCATGGGAACGAGTGTAAACCGTGGCTGTCTCCTGC
AGTGAGCCACCGCGGCAGGCACGTTGACTGTTTTACTGACATCACTCAAAAGCTAAAGCAATAACATTCTCCTGCGT
TGCTGAGTCAGCTGTTCATTTGTCCGCCAGCTCCTGGACTGGATGTGTGAAAGGCATCACATTTCCATTTTCCTCCG
TGTAAATGTTTTATGTGTTCGCCTACTGATCCCATTCGTTGCTTCTATTGTAAATATTTGTCATTTGTATTTATTAT
CTCTGTGTTTTCCCCCTAAGGCATAAAATGGTTTACTGTGTTCATTTGAACCCATTTACTGATCTCTGTTGTATATT
TTTCATGCCACTGCTTTGTTTTCTCCTCAGAAGTCGGGTAGATAGCATTTCTATCCCATCCCTCACGTTATTGGAAG
CATGCAACAGTATTTATTGCTCAGGGTCTTCTGCTTAAAACTGAGGAAGGTCCACATTCCTGCAAGCATTGATTGAG
ACATTTGCACAATCTAAAATGTAAGCAAAGTAGTCATTAAAAATACACCCTCTACTTGGGCTTTATACTGCATACAA
ATTTACTCATGAGCCTTCCTTTGAGGAAGGATGTGGATCTCCAAATAAAGATTTAGTGTTTATTTTGAGCTCTGCAT
CTTAACAAGATGATCTGAACACCTCTCCTTTGTATCAATAAATAGCCCTGTTATTCTGAAGTGAGAGGACCAAGTAT
AGTAAAATGCTGACATCTAAAACTAAATAAATAGAAAACACCAGGCCAGAACTATAGTCATACTCACACAAAGGGAG
AAATTTAAACTCGAACCAAGCAAAAGGCTTCACGGAAATAGCATGGAAAAACAATGCTTCCAGTGGCCACTTCCTAA
GGAGGAACAACCCCGTCTGATCTCAGAATTGGCACCACGTGAGCTTGCTAAGTGATAATATCTGTTTCTACTACGGA
TTTAGGCAACAGGACCTGTACATTGTCACATTGCATTATTTTTCTTCAAGCGTTAATAAAAGTTTTAAATAAATGGC
T

<210> SEQ ID NO 925

<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 925

>R20779_node_1

GGAGGAGGAGGGGAAGCGGCGAAGGAGGAAGAGGAGGA

<210> SEQ ID NO 926

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 926
>R20779_node_3
GAGCTCCCAGCAGAAAGGCCGCCTGTCCCTGCAGAATACAG

<210> SEQ ID NO 927

<211> Length : 11

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 927
>R20779_node_10
CTGCCCAGGAG

<210> SEQ ID NO 928

<211> Length : 53

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 928
>R20779_node_11
AACACCCGGGTGATAGTGGAGATGATCCATTTCAAGGACTTGCTGCTGCACGA

<210> SEQ ID NO 929

<211> Length : 73

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 929
>R20779_node_14
ATGCTACAAGATAGAAATTACTATGCCCAAGAGGAGGAAAGTGAAGCTAAGAGATTAGAGAACTCGGACTGAG

<210> SEQ ID NO 930

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 930
>R20779_node_17
ACCCTACGTGGACCTCGTGAACTTGCTGCTGAC

<210> SEQ ID NO 931

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 931
>R20779_node_19
GGCCCACCACGG

<210> SEQ ID NO 932

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 932
>R20779_node_20

GGAAGCAGGACATCACCTCCCAGAGCCCAG

<210> SEQ ID NO 933

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 933
>R20779_node_22
TCCGCGGAAGTCAGGGCGGCTGGATTCCAGGACAGGAGTGAATGTAAAAATAAATATCGCTTAGAATGCAGGAGAAG
GGTGGAGAGGAGGCAGGGGCCGAGGG

<210> SEQ ID NO 934

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 934
>R20779_node_23
GGTGCTTGGTGCCAAACTGAAATTCAGTTTCTTGTGTGGGGCCTTGCGGTTCAGAGCTCTTGGCGAGGGTGGAGGGA

<210> SEQ ID NO 935

<211> Length : 5

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 935
>R20779_node_25
CAGAA

<210> SEQ ID NO 936

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 936
>R20779_node_29
CTGAATTTCACTCACAG

Segment nucleic acid sequences:

<210> SEQ ID NO 937

<211> Length : 167

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 937
>R38144_PEA_2_node_21
TTCATGGCGTGAACCCAGGAGAGACCCCTGTCACCTGTACGGCAGGGATTGGGACCTTCATTGTTGAATTTGCCACC
CTGAGCAGCCTCACTGGTGACCCGGTGTTCGAAGATGTGGCCAGAGTGGCTTTGATGCGCCTCTGGGAGAGCCGGTC
AGATATCGGGCTG

<210> SEQ ID NO 938

<211> Length : 142

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 938
>R38144_PEA_2_node_26
GTCGGCAACCACATTGATGTGCTCACTGGCAAGTGGGTGGCCCAGGACGCAGGCATCGGGGCTGGCGTGGACTCCTA
CTTTGAGTACTTGGTGAAAGGAGCCATCCTGCTTCAGGATAAGAAGCTCATGGCCATGTTCCTAG

<210> SEQ ID NO 939

<211> Length : 125

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 939
>R38144_PEA_2_node_29
AGTATAACAAAGCCATCCGGAACTACACCCGCTTCGATGACTGGTACCTGTGGGTTCAGATGTACAAGGGGACTGTG
TCCATGCCAGTCTTCCAGTCCTTGGAGGCCTACTGGCCTGGTCTTCAG

<210> SEQ ID NO 940

<211> Length : 145

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 940
>R38144_PEA_2_node_31
AGCCTCATTGGAGACATTGACAATGCCATGAGGACCTTCCTCAACTACTACACTGTATGGAAGCAGTTTGGGGGGCT
CCCGGAATTCTACAACATTCCTCAGGGATACACAGTGGAGAAGCGAGAGGGCTACCCACTTCGGCCAG

<210> SEQ ID NO 941

<211> Length : 172

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 941
>R38144_PEA_2_node_46
AAGCTGGACAACCGCATGGAGTCGTTCTTCCTGGCCGAGACTGTGAAATACCTCTACCTCCTGTTTGACCCAACCAA
CTTCATCCACAACAATGGGTCCACCTTCGACGCGGTGATCACCCCCTATGGGGAGTGCATCCTGGGGGCTGGGGGGT
ACATCTTCAACACAGAAG

<210> SEQ ID NO 942

<211> Length : 375

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 942

>R38144_PEA_2_node_47

CTCACCCCATCGACCCTGCCGCCCTGCACTGCTGCCAGAGGCTGAAGGAAGAGCAGTGGGAGGTGGAGGACTTGATG

AGGGAATTCTACTCTCTCAAACGGAGCAGGTCGAAATTTCAGAAAAACACTGTTAGTTCGGGGCCATGGGAACCTCC

AGCAAGGCCAGGAACACTCTTCTCACCAGAAAACCATGACCAGGCAAGGGAGAGGAAGCCTGCCAAACAGAAGGTCC

CACTTCTCAGCTGCCCCAGTCAGCCCTTCACCTCCAAGTTGGCATTACTGGGACAGGTTTTCCTAGACTCCTCATAA

CCACTGGATAATTTTTTTATTTTTATTTTTTTGAGGCTAAACTATAATAAATTGCTTTTGGCTATCA

<210> SEQ ID NO 943

<211> Length : 122

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 943

>R38144_PEA_2_node_49

AAAAGATCTCGCTCTGTTGCCCAGGCTGGAGTGCAGTGGTGTGATCACGACTCACCGCAGCCTTGACCTCCCACACT

CAAGCAATCCTCCTGCCTTAGCCTTCCAAGTAGCTGGAACTCCAG

<210> SEQ ID NO 944

<211> Length : 105

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 944

>R38144_PEA_2_node_0

GGATTCCCGGAAGAACCCGCAGCAGCTCCCAGGATGAACTGGTTGCAGTGGCTGCTGCTGCTGCGGGGGCGCTGAGA
GGACACGAGCTCTATGCCTTTCCGGCTG

<210> SEQ ID NO 945

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 945
>R38144_PEA_2_node_1
CTCATCCCGCTCGGCCTCCTGTGCGCGCTGCTGCCTCAGCACCATGGTGCGCCAGGTCCCGACGGCTCCGCGCCAGA
TCCCGCCCACTACAG

<210> SEQ ID NO 946

<211> Length : 102

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 946
>R38144_PEA_2_node_4
GGAGCGAGTCAAGGCCATGTTCTACCACGCCTACGACAGCTACCTGGAGAATGCCTTTCCCTTCGATGAGCTGCGAC
CTCTCACCTGTGACGGGCACGACAC

<210> SEQ ID NO 947

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 947

>R38144_PEA_2_node_5
CTGGGGCAG

<210> SEQ ID NO 948

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 948
>R38144_PEA_2_node_7
TTTTTCTCTGACTCTAATTGATGCACTGGACACCTTGCTG

<210> SEQ ID NO 949

<211> Length : 106

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 949
>R38144_PEA_2_node_11
ATTTTGGGGAATGTCTCAGAATTCCAAAGAGTGGTTGAAGTGCTCCAGGACAGCGTGGACTTTGATATTGATGTGAA
CGCCTCTGTGTTTGAAACAAACATTCGAG

<210> SEQ ID NO 950

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 950
>R38144_PEA_2_node_14
TGGTAG

<210> SEQ ID NO 951

<211> Length : 27

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 951
>R38144_PEA_2_node_15
GAGGACTCCTGTCTGCTCATCTGCTCT

<210> SEQ ID NO 952

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 952
>R38144_PEA_2_node_16
CCAAGAAGGCTGGGGTGGAAGTAGAGGCTGGATGGCCCTGTTCCGGGCCTCTCCTGAGAATGGCTGAGGAGGCGGCC
CGAAAACTCCTCCCAG

<210> SEQ ID NO 953

<211> Length : 35

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 953
>R38144_PEA_2_node_19
CCTTTCAGACCCCCACTGGCATGCCATATGGAACA

<210> SEQ ID NO 954

<211> Length : 10

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 954
>R38144_PEA_2_node_20
GTGAACTTAC

<210> SEQ ID NO 955

<211> Length : 89

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 955
>R38144_PEA_2_node_36
AACTTATTGAAAGCGCAATGTACCTCTACCGTGCCACGGGGGATCCCACCCTCCTAGAACTCGGAAGAGATGCTGTG
GAATCCATTGAA

<210> SEQ ID NO 956

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 956
>R38144_PEA_2_node_37
AAAATCAGCAAGGTGGAGTGCGGATTTGCAACA

<210> SEQ ID NO 957

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 957
>R38144_PEA_2_node_43
CTTGCTTCCTTCTCCCACAT

<210> SEQ ID NO 958

<211> Length : 5

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 958
>R38144_PEA_2_node_44
GTCAG

<210> SEQ ID NO 959

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 959
>R38144_PEA_2_node_45
ATCAAAGATCTGCGAGACCAC

<210> SEQ ID NO 960

<211> Length : 74

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 960
>R38144_PEA_2_node_51
GTGGTGGTTAATTTTATGTGTCAACCTGGCTGGACCACTGGGTACTCCGATATTTGGTCAAACATTATTCTGAG

Segment nucleic acid sequences:

<210> SEQ ID NO 961

<211> Length : 184

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 961
>HUMOSTRO_PEA_1_PEA_1_node_0
GTGGCAGAAAACCTCATGACACAATCTCTCCGCCTCCCTGTGTTGGTGGAGGATGTCTGCAGCAGCATTTAAATTCT
GGGAGGGCTTGGTTGTCAGCAGCAGCAGGAGGAGGCAGAGCACAGCATCGTCGGGACCAGACTCGTCTCAGGCCAGT
TGCAGCCTTCTCAGCCAAACGCCGACCAAG

<210> SEQ ID NO 962

<211> Length : 189

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 962
>HUMOSTRO_PEA_1_PEA_1_node_10
CACTAAAGATGTACCTACCCCTCCACAACAGATGAAACTGTGCCAGCCAAACAACAAATGGGCATTGTCCCCAGAAG
CTTGGACAAAAAGGCACACAGAGTTCAATTCCAGTTGAACAGAATAAAGGCCAAAATAGAGCTGCCTTGGGGGTCAC
TGCAATTAGACTGCTTAATGAAGACATTAAAAGAA

<210> SEQ ID NO 963
<211> Length : 266
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 963
>HUMOSTRO_PEA_1_PEA_1_node_16
GTATTTTTAAACTTCTCATAATTAAACTACAGTGATGAAAGATAGCCACACTCAGGCCATTTGGGCTGCTCAGATGA
ATCCTGCCTGCCTGCTGGCAAACATGTGCTTAGGACATTGACTGATCTGCCATGTTGGCTTCTCTCTGTGTTAAGCC
ATCCACAGATGAGGCTGAAAAATAAAAACTGCTTTGGATTAAAAAGGTTAACTTTTGAATAAAAAAGCTAGGCATGT
GTGATGCGCACTAACACGTGCCATTCCTTCTTCAG

<210> SEQ ID NO 964
<211> Length : 164
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 964
>HUMOSTRO_PEA_1_PEA_1_node_23
ACTGATGATTCTCACCAGTCTGATGAGTCTCACCATTCTGATGAATCTGATGAACTGGTCACTGATTTTCCCACGGA
CCTGCCAGCAACCGAAGTTTTCACTCCAGTTGTCCCCACAGTAGACACATATGATGGCCGAGGTGATAGTGTGGTTT
ATGGACTGAG

<210> SEQ ID NO 965

<211> Length : 230

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 965

>HUMOSTRO_PEA_1_PEA_1_node_31

AGTGCTGAAACCCACAGCCACAAGCAGTCCAGATTATATAAGCGGAAAGCCAATGATGAGAGCAATGAGCATTCCGA

TGTGATTGATAGTCAGGAACTTTCCAAAGTCAGCCGTGAATTCCACAGCCATGAATTTCACAGCCATGAAGATATGC

TGGTTGTAGACCCCAAAAGTAAGGAAGAAGATAAACACCTGAAATTTCGTATTTCTCATGAATTAGATAGTGCATC

<210> SEQ ID NO 966

<211> Length : 136

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 966

>HUMOSTRO_PEA_1_PEA_1_node_43

TGGTGTGAATAAATCTTTTATCTTGAATGTAATAAGAATTTGGTGGTGTCAATTGCTTATTTGTTTTCCCACGGTTG

TCCAGCAATTAATAAAACATAACCTTTTTTACTGCCTATATAATGTTTTTAAAGGTTTA

<210> SEQ ID NO 967

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 967

>HUMOSTRO_PEA_1_PEA_1_node_3

GAAAACTCACTACCATGAGAATTGCA

<210> SEQ ID NO 968

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 968
>HUMOSTRO_PEA_1_PEA_1_node_5
GTGATTTGCTTTTGCCTCCTAGGCATCACCTGTGCCATACCA

<210> SEQ ID NO 969

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 969
>HUMOSTRO_PEA_1_PEA_1_node_7
GTTAAACAGGCTGATTCTGGAAGTTCTGAGGAAAAGCAG

<210> SEQ ID NO 970

<211> Length : 87

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 970
>HUMOSTRO_PEA_1_PEA_1_node_8
GTAAGCATCTTTTATGTTTTTATATAGTTAAATCATTTACTCAATTATGGCGAGAGGTGCAAGAAACGTATTTGCTG
CGATATTACT

<210> SEQ ID NO 971

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 971
>HUMOSTRO_PEA_1_PEA_1_node_15
CTTTACAACAAATACCCAGATGCTGTGGCCACATGGCTAAACCCTGACCCATCTCAGAAGCAGAATCTCCTAGCCCC
ACAG

<210> SEQ ID NO 972

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 972
>HUMOSTRO_PEA_1_PEA_1_node_17
AATGCTGTGTCCTCTGAAGAAACCAATGACTTTAAACAAGAG

<210> SEQ ID NO 973

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 973
>HUMOSTRO_PEA_1_PEA_1_node_20

ACCCTTCC

<210> SEQ ID NO 974

<211> Length : 50

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 974
>HUMOSTRO_PEA_1_PEA_1_node_21
AAGTAAGTCCAACGAAAGCCATGACCACATGGATGATATGGATGATGAAG

<210> SEQ ID NO 975

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 975
>HUMOSTRO_PEA_1_PEA_1_node_22
ATGATGATGACCATGTGGACAGCCAGGACTCCATTGACTCGAACGACTCTGATGATGTAGATGAC

<210> SEQ ID NO 976

<211> Length : 37

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 976
>HUMOSTRO_PEA_1_PEA_1_node_24

GTCAAAATCTAAGAAGTTTCGCAGACCTGACATCCAG

<210> SEQ ID NO 977

<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 977
>HUMOSTRO_PEA_1_PEA_1_node_26
TACCCTGATGCTACAGAC

<210> SEQ ID NO 978

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 978
>HUMOSTRO_PEA_1_PEA_1_node_27
GAGGACATCACCTCACACATGGAAAG

<210> SEQ ID NO 979

<211> Length : 52

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 979

>HUMOSTRO_PEA_1_PEA_1_node_28
CGAGGAGTTGAATGGTGCATACAAGGCCATCCCCGTTGCCCAGGACCTGAAC

<210> SEQ ID NO 980

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 980
>HUMOSTRO_PEA_1_PEA_1_node_29
GCGCCTTCTGATTGGGACAGCCGTGGGAAGGACAGTTATGAAACGAGTCAG

<210> SEQ ID NO 981

<211> Length : 12

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 981
>HUMOSTRO_PEA_1_PEA_1_node_30
CTGGATGACCAG

<210> SEQ ID NO 982

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 982
>HUMOSTRO_PEA_1_PEA_1_node_32
TTCTGAGGTCAATTAAAAGGAGAAAAAATACAAT

<210> SEQ ID NO 983

<211> Length : 41

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 983
>HUMOSTRO_PEA_1_PEA_1_node_34
TTCTCACTTTGCATTTAGTCAAAAGAAAAAATGCTTTATAG

<210> SEQ ID NO 984

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 984
>HUMOSTRO_PEA_1_PEA_1_node_36
CAAAATGAAAGAGAACATGAAATGCTTCTTTCTCAG

<210> SEQ ID NO 985

<211> Length : 119

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 985
>HUMOSTRO_PEA_1_PEA_1_node_37
TTTATTGGTTGAATGTGTATCTATTTGAGTCTGGAAATAACTAATGTGTTTGATAATTAGTTTAGTTTGTGGCTTCA
TGGAAACTCCCTGTAAACTAAAAGCTTCAGGGTTATGTCTAT

<210> SEQ ID NO 986
<211> Length : 11
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 986
>HUMOSTRO_PEA_1_PEA_1_node_38
GTTCATTCTAT

<210> SEQ ID NO 987
<211> Length : 91
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 987
>HUMOSTRO_PEA_1_PEA_1_node_39
AGAAGAAATGCAAACTATCACTGTATTTTAATATTTGTTATTCTCTCATGAATAGAAATTTATGTAGAAGCAAACAA
AATACTTTTACCCA

<210> SEQ ID NO 988
<211> Length : 18

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 988
>HUMOSTRO_PEA_1_PEA_1_node_40
CTTAAAAAGAGAATATAA

<210> SEQ ID NO 989

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 989
>HUMOSTRO_PEA_1_PEA_1_node_41
CATTTT

<210> SEQ ID NO 990

<211> Length : 60

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 990
>HUMOSTRO_PEA_1_PEA_1_node_42
ATGTCACTATAATCTTTTGTTTTTTAAGTTAGTGTATATTTTGTTGTGATTATCTTTTTG

Segment nucleic acid sequences:

<210> SEQ ID NO 991

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 991

>R11723_PEA_1_node_13

ACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCAGTTGACCACTTAGTTCTCAAGAAGC

AACTATCTCTTTCATGTGCCTTCTGAGGAAGTATTCAGAGGGGGAATATCAAATGTCTTTCCCTTGGACTCTCCCA

<210> SEQ ID NO 992

<211> Length : 744

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 992

>R11723_PEA_1_node_16

GGTCTCACTGTGTCACGAGGCTGGAGTGCAGTGGAACAATTTCAGCACACTGCAACCTCTGCCTCCCAGGCTCAAAT

GATCATCCCACCTAAGCCTCCGGAGTAGCTGGGACCACAGGCAAGCGCCACCATGCCCAGCTGATACCAATGTCTTT

TAAAAAATGTTGTATGTGGAAATAAATTGAGACTTATAGAAAAGCTGCAAAAATAGTGCAGTTTCTATATATCCTTC

CCCCATCTTTGGCTAGTGTTAACAATCTACATAACCGCAGTACGATGATCAAGGCTAGGAAATTAACATTGGCACAG

TACTGTTAATGAAACCATGCTTTGTTTTGAGATTCCCACAGTTTTGCCTTTTTCTGTTCCAAGATCCTATCCAGGAT

CCCACGTTGCATTTCATTGTCATGTCTCCTTCTCCTCTAACCTCTGACAATGCATCATTCTTTCCATGTCTTTTGTG

ATGTTGACACTTTTGAAGAGGACTGGTCCAGATTTTTGTACACTGTCCCTCAGTTTGGGATTGTCTGCTGTTTTCTC

ATGAACAGATAGAGGTTTTGCATTTTTGACAAGAATCCTCAGAAGAGATGCACCCTTCTCAGTGCACTGTAGCAAGG

GGCGCATGCTGTCAATGTCTTACTGGTGATGTTAACTTTGATCGCTTTTGATTCAGATAGTATCTGCTGGGTTTTTC

CACTGTAAAGTTACTATTTTTTCCATTGTAATTAATAAATAACTTGAGGGA

<210> SEQ ID NO 993

<211> Length : 174

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 993

>R11723_PEA_1_node_19

GTACCAGTCCTTCTGCTCCCCAGGGAAACTGAACTCAGTTTGCATCAGCTGCTGCAACACCCCTCTTTGTAACGGGC
CAAGGCCCAAGAAAAGGGGAAGTTCTGCCTCGGCCCTCAGGCCAGGGCTCCGCACCACCATCCTGTTCCTCAAATTA
GCCCTCTTCTCGGCACACTG

<210> SEQ ID NO 994

<211> Length : 309

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 994

>R11723_PEA_1_node_2

AGAAGAGGAAGACAGGAAGGGGGTGGGGATGTGAAGCGACCGTCCCAGCCTTCCCCGCCCGCCACCCCCACCCCAAC
TCGGCAGCCGTCACGTGATGCCTGGAGTGGGAGGTGGGGAGAAAAGGCGAGACTTTTGTGGGTGCTCCCGATCGCCA
GTAGTTCCTTCAGTCTCAGCCGCCAACTCCGGAGGCGCGGTGCTCGGCCCGGGAGCGCGAGCGGGAGGAGCAGAGAC
CCGCAGCCGGGAGCCCGAGCGCGGGCGATGCAGGCTCCGCGAGCGGCACCTGCGGCTCCTCTAAGCTACGACCGTCG
T

<210> SEQ ID NO 995

<211> Length : 487

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 995

>R11723_PEA_1_node_22
GTGAGTTTCTTCTGGGTGTCCTTTTATTCTGGGTAGGGAGCGGGAGTCCGTGTTCTCTTTTGTTCCTGTGCAAATAA
TGAAAGAGCTCGGTAAAGCATTCTGAATAAATTCAGCCTGACTGAATTTTCAGTATGTACTTGAAGGAAGGAGGTGG
AGTGAAAGTTCACCCCCATGTCTGTGTAACCGGAGTCAAGGCCAGGCTGGCAGAGTCAGTCCTTAGAAGTCACTGAG
GTGGGCATCTGCCTTTTGTAAAGCCTCCAGTGTCCATTCCATCCCTGATGGGGGCATAGTTTGAGACTGCAGAGTGA
GAGTGACGTTTTCTTAGGGCTGGAGGGCCAGTTCCCACTCAAGGCTCCCTCGCTTGACATTCAAACTTCATGCTCCT
GAAAACCATTCTCTGCAGCAGAATTGGCTGGTTTCGCGCCTGAGTTGGGCTCTAGTGACTCGAGACTCAATGACTGG
GACTTAGACTGGGGCTCGGCCTCGC

<210> SEQ ID NO 996

<211> Length : 418

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 996
>R11723_PEA_1_node_31
GGAGAGACAGAGAAAAGAAAAACACAGCATGAGAACACAGTAAATAAATAAAACCATAAAATATTTAGCCCCTCTGT
TCTGTGCTTACTGGCCAGGAAATGGTACCAATTTTTCAGTGTTGGACTTGACAGCTTCTTTTGCCACAAGCAAGAGA
GAATTTAACACTGTTTCAAACCCGGGGGAGTTGGCTGTGTTAAAGAAAGACCATTAAATGCTTTAGACAGTGTATTT
ATACCAGTTGATGTCTGTTAATTTTAAAAAAATGTTTTCATTGGTGTTTGTTTGCGTATCCAGAAAGCAGTTCATGT
TATCCATAAATCTGGTTTTGTCTTTTTTTGTTTTAAAGAAAAAGATGTATACATACAGTATAGCTGCATTAGATAAA
GCAGTGTTTGTATTTTAAAGGATGTCTGCACAA

<210> SEQ ID NO 997

<211> Length : 44

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 997
>R11723_PEA_1_node_10
GCTTTGCGCTGCAAATCCAGTGCTACCAGTGTGAAGAATTCCAG

<210> SEQ ID NO 998

<211> Length : 94

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 998
>R11723_PEA_1_node_11
CTGAACAACGACTGCTCCTCCCCCGAGTTCATTGTGAATTGCACGGTGAACGTTCAAGACATGTGTCAGAAAGAAGT
GATGGAGCAAAGTGCCG

<210> SEQ ID NO 999

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 999
>R11723_PEA_1_node_15
ACAG

<210> SEQ ID NO 1000

<211> Length : 58

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1000
>R11723_PEA_1_node_18
GGATCATGTACCGCAAGTCCTGTGCATCATCAGCGGCCTGTCTCATCGCCTCTGCCGG

<210> SEQ ID NO 1001

<211> Length : 11

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1001
>R11723_PEA_1_node_20
CTGAAGCTGAA

<210> SEQ ID NO 1002

<211> Length : 63

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1002
>R11723_PEA_1_node_21
GGAGATGCCACCCCCTCCTGCATTGTTCTTCCAGCCCTCGCCCCCAACCCCCCACCTCCCTGA

<210> SEQ ID NO 1003

<211> Length : 30

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1003
>R11723_PEA_1_node_23
TCTGAAAAGTGCTTAAGAAAATCTTCTCAG

<210> SEQ ID NO 1004

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1004
>R11723_PEA_1_node_24
TTCTCCTTGCAGAGGACTGGCGCCGGGACGCGAAGAGCAACGGGCGCTGCACAAAGCGGGCGCTGTCGGTGGTGGAG
TGCGCAT

<210> SEQ ID NO 1005

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1005
>R11723_PEA_1_node_25
GTACGCGCAGGCGCTTCTCGTGGTTG

<210> SEQ ID NO 1006

<211> Length : 113

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1006
>R11723_PEA_1_node_26

GCGTGCTGCAGCGACAGGCGGCAGCACAGCACCTGCACGAACACCCGCCGAAACTGCTGCGAGGACACCGTGTACAG
GAGCGGGTTGATGACCGAGCTGAGGTAGAAAAACGT

<210> SEQ ID NO 1007

<211> Length : 82

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1007
>R11723_PEA_1_node_27
CTCCGAGAAGGGGAGGAGGATCATGTACGCCCGGAAGTAGGACCTCGTCCAGTCGTGCTTGGGTTTGGCCGCAGCCA
TGATC

<210> SEQ ID NO 1008

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1008
>R11723_PEA_1_node_28
CTCCGAATCTGGTTGGGCATCCAG

<210> SEQ ID NO 1009

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1009
>R11723_PEA_1_node_29
CATACGGCCAATGTCACAACAATCAGCC

<210> SEQ ID NO 1010

<211> Length : 10

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1010
>R11723_PEA_1_node_3
CTCCGCGGCA

<210> SEQ ID NO 1011

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1011
>R11723_PEA_1_node_30
CTGGGCAGACACGAGCAGGAG

<210> SEQ ID NO 1012

<211> Length : 52

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1012
>R11723_PEA_1_node_4

GCAGCGCGGGCCCCAGCAGCCTCGGCAGCCACAGCCGCTGCAGCCGGGGCAG

<210> SEQ ID NO 1013

<211> Length : 43

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1013
>R11723_PEA_1_node_5
CCTCCGCTGCTGTCGCCTCCTCTGATGCGCTTGCCCTCTCCCG

<210> SEQ ID NO 1014

<211> Length : 32

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1014
>R11723_PEA_1_node_6
GCCCCGGGACTCCGGGAGAATGTGGGTCCTAG

<210> SEQ ID NO 1015

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1015

>R11723_PEA_1_node_7
GCATCGCGGCAACTTTTTGCGGATTGTTCTTGCTTCCAG

<210> SEQ ID NO 1016

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1016
>R11723_PEA_1_node_8
GTGAGAATACCCAGAGGCCAGCAGCCGAGGCCAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1017

<211> Length : 438

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1017
>R16276_PEA_1_node_0
GTGGAGGAGGATGGTGGGGAGTGGTGGTGTCTTCGTCCTGGGAGAAGGCGAAGCAACTTCCAGGAGGAAACGGGCGT
TTCCTTCCCACGCGCTCGAGCGAGCCCTGGGTCCTGGCCTCGGAACTCCACCCAGCCCCTCCCCACCCTCTGGGAAA
AGCCAGTCGCCACACACAGGCACACGCAGGCCCCGGCGCCGCGCCCTAAGGAGAGCAGCACCCACAGCCAATTGCCA
TGGCAACCCCGGGGTTCGTTCCACTTCCCCACCCAGCCGATCTCCCCCCTCCTCCCTGCACTGCAGCCAACCGGCTT
GTGCGCGTCCCAGGAGCGCGCTATAAAACCTGTGCTGGGCGTGATCGGCAAGCACCGGACCAGGGGGAAGGCGAGCA
GTGCCAATCTACAGCGAAGAAAGTCTCGTTTGGTAAAAGCGAGAAGGGAAAGC

<210> SEQ ID NO 1018

<211> Length : 122

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1018

>R16276_PEA_1_node_6

GTAATCCTGCTCCCTCTGCTGTTTGACCTCTTCTCCTGCAGCTAAGTGAAGCTGCTTCCTCCCTTCTCTTTTGTATT

CCCCTTCCCAGAGGGCGATAAGCAAATAATAATAATGCAATAAAT

<210> SEQ ID NO 1019

<211> Length : 90

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1019

>R16276_PEA_1_node_1

CTGAGCATGCAGAGTGTGCAGAGCACGAGCTTTTGTCTCCGAAAGCAGTGCCTTTGCCTGACCTTCCTGCTTCTCCA

TCTCCTGGGACAG

<210> SEQ ID NO 1020

<211> Length : 111

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1020

>R16276_PEA_1_node_4
GTCGCTGCGACTCAGCGCTGCCCTCCCCAGTGCCCGGGCCAGTGCCCTGCGACGCCGCCGACCTGCGCCCCCGGGGT
GCGCGCGGTGCTGGACGGCTGCTCATGCTGTCTG

<210> SEQ ID NO 1021

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1021
>R16276_PEA_1_node_5
GTGTGTGCCCGCCAGCGTGGCGAGAGCTGCTCAGATCTGGAGCCATGCGACGAGAGCAGTGGCCTCTACTGTGATCG
CAGCGCGGACCCCAGCAACCAGACTGGCATCTGCACGG

Segment nucleic acid sequences:

<210> SEQ ID NO 1022

<211> Length : 232

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1022
>H61775_node_2
ATCTGGTGGTTCTCCGGAGAGCAGCTTCCTTGGGTGTTACATGAGCCAAGCCCTCACTGTACAGAAGAGTGAGAGCT
GAAACCTGTTCCCTGAGCTGATCAGAAGGACATCCCTTGGCCCCTCCATCTGGGCTCCTGTGGATAGGAGGGGCTGG

GTGAGCAGGCCAGCTGGGCTATGGTGTGGTGCCTCGGCCTGGCCGTCCTCAGCCTGGTCATCAGCCAGGGGGCTGAC
G

<210> SEQ ID NO 1023

<211> Length : 189

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1023
>H61775_node_4
GTCGAGGGAAGCCTGAGGTGGTATCGGTGGTGGGCCGGGCTGGGGAGAGTGTGGTGCTGGGCTGTGACCTGCTGCCC
CCGGCCGGCCGGCCCCCCCTGCATGTCATCGAGTGGCTGCGCTTTGGATTCCTGCTTCCCATCTTCATCCAGTTCGG
CCTCTACTCTCCCCGAATTGACCCTGATTACGTGG

<210> SEQ ID NO 1024

<211> Length : 201

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1024
>H61775_node_6
GTTCTCCTCAGGTGGGAGGTAGGGAGGTATCAGCAAGAAAGGTGGGCTGGGTAGAGTCGCACAAGGCCTCCTATGAA
CGGCTTTGTCCCTGCTCTGATCTCATCTCCAGCTCTGCTGCCTTAACTCTGCTTAATAAGCATGGCTGTGCTCCCAA
GCAGTGTTAATTCATTGAAAGATGTCATTCATTTACACACACACACA

<210> SEQ ID NO 1025

<211> Length : 698

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1025
>H61775_node_8
GTGACTGTGGGTTCCCTGCCTTCCGAGAGCTTAAGAGAGCAGAGACTGTGTCTCCTGTTTTCTTCACACGCCGCTGC
ATATGGGAAGATCTGAAGTCAACAGGCTTTAGCCCTGCAGGTGGAGGGAGGCCTCCAGGAGGTGGGCCCAGGACTCA
GGAGGACTCAGGGCTGCCCTGCTGGCGATCTTCCTGTTCTGTAACACTACAGGTCTAGCAGTCCAGCTGTCACAGAA
AAGCTAGGACATGCAGTATGCTTCTTTGGATATTCTGAGTAACATTTGGACTGTTACCCATTGGCTACCAGCATCTC
CCAAGTGAGAATACATAGATTACCCCCAGTGCCCTGAACAGCACTCGGTCCTAACACCCGTGTCCATGGAAAGCACG
CCGCGTCTGGAGAAAGAAGCCGAAGGCTCTTGTCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATT
CCTTCAGCTACAATGGAATTCTTGGGAGGATTAAATATGGTGACAACGCCTAATATTAGATGGCCTGTATTCCACAC
TCAATCTTCCTTCCCTCTTCTTCCTTCTTTGTAGAGCTATAATGAAAAGTATCATGTGGGACACAGAAGAGGTTGCA
GTCTGGGGTCTGCAGGGCTTAGCGGCCAGGCAGATTAGCTTTCTTGAGGAATCCTGACAGTGGGTGGAAGGGTATGA
TGATG

<210> SEQ ID NO 1026

<211> Length : 86

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1026
>H61775_node_0
GGAGGCGCTCGGGGCATCCGAGGCGGGGAGGCGGGTCCGCCCCCTATTGTGTAGCGGCGAGAGTGGAGCCGAGCGGT
GCGGAGCAG

<210> SEQ ID NO 1027

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1027
>H61775_node_5
GATAAGA

Segment nucleic acid sequences:

<210> SEQ ID NO 1028

<211> Length : 203

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1028
>M85491_PEA_1_node_0
TCTGCTGGCTGCGCGGTGGCGGCGGCTGTGTGTGCGCCGCGCCTTGCCGCCCCCCCTGGCCCCCCGAGCCCGGGGCG
CGCGCTCCCGCCCGGGCCGTCCGGGCCCCGCGGCGCCGCGGCCCGAGGCCCCGGGAAGCGCAGCCATGGCTCTGCGG
AGGCTGGGGGCCGCGCTGCTGCTGCTGCCGCTGCTCGCCGCCGTGGAAG

<210> SEQ ID NO 1029

<211> Length : 229

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1029
>M85491_PEA_1_node_13
AGAGACAGGATCTCACTATGTTGTCCAGGCTGGTCTTGAACTCGTGGCCACAAATGATCCTCCCACCTCAGCCTCCC
AAAGTGTTGGAATTATAGGCATGAACCACCATGCCCAGGAGGAGAATTTTTGATAATAATATTTTGTGGACATCTTT
GCATATCATGTCAGAGCTATAACATCATTGTGGAGAAGCTCTTAGGATCCCATAGAATAAATGTACCGTAATTTA

<210> SEQ ID NO 1030

<211> Length : 336

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1030

>M85491_PEA_1_node_21

CCATCCCCTCCGCGCCCCAGGCTGTGATTTCCAGTGTCAATGAGACCTCCCTCATGCTGGAGTGGACCCCTCCCCGC

GACTCCGGAGGCCGAGAGGACCTCGTCTACAACATCATCTGCAAGAGCTGTGGCTCGGGCCGGGGTGCCTGCACCCG

CTGCGGGGACAATGTACAGTACGCACCACGCCAGCTAGGCCTGACCGAGCCACGCATTTACATCAGTGACCTGCTGG

CCCACACCCAGTACACCTTCGAGATCCAGGCTGTGAACGGCGTTACTGACCAGAGCCCCTTCTCGCCTCAGTTCGCC

TCTGTGAACATCACCACCAACCAGGCAG

<210> SEQ ID NO 1031

<211> Length : 125

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1031

>M85491_PEA_1_node_23

CTCCATCGGCAGTGTCCATCATGCATCAGGTGAGCCGCACCGTGGACAGCATTACCCTGTCGTGGTCCCAGCCAGAC

CAGCCCAATGGCGTGATCCTGGACTATGAGCTGCAGTACTATGAGAAG

<210> SEQ ID NO 1032

<211> Length : 1,305

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1032

>M85491_PEA_1_node_24

GTACCTATTGGCTGGGTGCTGTCCCCATCACCCACCTCCCTGAGGGCCCCTCTCCCAGGCTGAGGCCTGGGAGTTCT
GCCCCACCGCAAGATGAGACGCACTGGTGCAGCAGAAAGAGCACTGGCCTTGGAGTCAGGCTGCCTGGCTCCCAATC
CAGCTCCGCTCCTTCCCACTGTGAGACCTCAGGCAGGTGCCTTGACCTCTCTGGATCTCACTTTTCTGGTCTGGAGG
ATACACCCAGCAATCTCAGTGAAATGCAACAGTCACATCCCTTTCCCTACCACGACCCTTTCATCTTGACCTCAGTG
GCTTGATGTTGGGAAAAACTGGGTTTCCAAAAAGCTGCACTTATGAAGTGATAATTAGTCACTCACCTCTTCTTCGA
CAGAGATTTGAAACAGCTCAAGAGAGCTTCCGCCTGCCCTGCTCTGAGTCCTGCTAAAACACCCACTTTCACTCGCC
TGCATGCCCTTTGCATGGGGAGAGGTGATTTCACTTTGAGCTTTTAAATCAGACCTTAATTACTCCCTTTGGGTGGA
AGCCCCTGGGATGGTAGAAGGATCACTGGACTAAGAGTGAGAAGCCGTAGGTTCAAATCCCAGCTCCGTCCTTCACC
AGCTATGTGACCTTGGGCAGGCGTCTTTCTCCCTCTGAACCTCAGTTTCCACCTGTGTCGAGTGTGGGTGAGACCCC
TCGCGGGGAGCTATGCAGGTTACGGAGAAAAGGCAGCACAGCACCCAGAATGGGACCTGGCCCTCAGCAGAGGCCAT
GTGTGTCCCTGGCCTTCCTCCTCTGCCCTGCCTGCTGCACAGTGGGCAATGGTGACAGGATGGGAGGCCAAGTGGAT
GTGGGGTCTGCACAGTACAGGGGCCAGGAGGTAGACAGCACAATTGCCCACCCACATGGCTGGACATCAGAGGCCCC
AGGAAGCCTCTCCTTTGAATGATCACTTCTCTTACCTGCTCCAGGAGGCAACAAACAGCCACAGAGGCTGCAAGGGC
ACCTGGGAAAGGCATCGCGGGGCTTCCATTCAGACTAGGTGTCAATGACTGACAGGGAGGCCTTTGGTTGAGGGCAA
GCCCACGGGGAACTGCAGATGGATGGAAGGGCTCTCCCTGAAGGCTGAGAGGAAGAGTGCAGTCAATTGCAGCCAGT
CCTGCTGGAGCCCAACTTTCTAGAGCCCAGCCCGGCCTTCCCACTCTGTTAACTGCTGGATCGGCTAACCAGGCCGG
TCTCCAGGGCCTTTCAAACACTTACCCAGCCTTTGCCGGCCGTCTTACCATTGCTTGCGTGCGTGTTCATCCC

<210> SEQ ID NO 1033

<211> Length : 404

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1033

>M85491_PEA_1_node_8

TGGGAAGAGGTGAGTGGCTACGATGAGAACATGAACACGATCCGCACGTACCAGGTGTGCAACGTGTTTGAGTCAAG
CCAGAACAACTGGCTACGGACCAAGTTTATCCGGCGCCGTGGCGCCCACCGCATCCACGTGGAGATGAAGTTTTCGG
TGCGTGACTGCAGCAGCATCCCCAGCGTGCCTGGCTCCTGCAAGGAGACCTTCAACCTCTATTACTATGAGGCTGAC

TTTGACTCGGCCACCAAGACCTTCCCCAACTGGATGGAGAATCCATGGGTGAAGGTGGATACCATTGCAGCCGACGA
GAGCTTCTCCCAGGTGGACCTGGGTGGCCGCGTCATGAAAATCAACACCGAGGTGCGGAGCTTCGGACCTGTGTCCC
GCAGCGGCTTCTACCTGGC

<210> SEQ ID NO 1034

<211> Length : 184

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1034
>M85491_PEA_1_node_9
CTTCCAGGACTATGGCGGCTGCATGTCCCTCATCGCCGTGCGTGTCTTCTACCGCAAGTGCCCCCGCATCATCCAGA
ATGGCGCCATCTTCCAGGAAACCCTGTCGGGGGCTGAGAGCACATCGCTGGTGGCTGCCCGGGGCAGCTGCATCGCC
AATGCGGAAGAGGTGGATGTACCCATCAAG

<210> SEQ ID NO 1035

<211> Length : 97

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1035
>M85491_PEA_1_node_10
CTCTACTGTAACGGGGACGGCGAGTGGCTGGTGCCCATCGGGCGCTGCATGTGCAAAGCAGGCTTCGAGGCCGTTGA
GAATGGCACCGTCTGCCGAG

<210> SEQ ID NO 1036

<211> Length : 91

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1036
>M85491_PEA_1_node_18
GTTGTCCATCTGGGACTTTCAAGGCCAACCAAGGGGATGAGGCCTGTACCCACTGTCCCATCAACAGCCGGACCACT
TCTGAAGGGGCCAC

<210> SEQ ID NO 1037

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1037
>M85491_PEA_1_node_19
CAACTGTGTCTGCCGCAATGGCTACTACAGAGCAGACCTGGACCCCCTGGACATGCCCTGCACAA

<210> SEQ ID NO 1038

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1038
>M85491_PEA_1_node_6
AAACGCTAATGGACTCCACTACAGCGACTGCTGAGCTGGGCTGGATGGTGCATCCTCCATCAGGG

Segment nucleic acid sequences:

<210> SEQ ID NO 1039

<211> Length : 810

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1039

>T39971_node_0

GAGACTGAGCCTGGGGACAGGGAGTGGCCTGCTCAGAAAAGACTCAGAAATTAAATCCAGTCCAGTGGGTTGATATT
TACCCAAATTTCCAGCCTGGGGAGATTGATGCACCCAAGAGAAGAACCCAGAAATGAAACTTTGTTCTTTTATGCTA
AAAAATAAAATTCCCCAGAGTGCTTACAATCTCTCCTCCCACTCCCTTTTTCCTGCCCTAAATAAATAATGGCGAAT
GAGCACCCAGCCAGGGATGTGTCTGATCAAACAATCATGGATCAATAGCTATGTTTGGAGAAGGAATTTGTGGCTGC
TCCAGCTACTGGGCATTTTGTCTGGTCCAGTTCATGTAATCTCCCAACACCCCATGAAGCAAGGCTTTGTTAATCCT
ATTTTACTGAAAATGAACTAAGACTCAGAGAGATAAAGCTGTTGCCCAATGAGCCTTCTTTCTGCCCTCCAGATCCA
CGGTGCTAATTCCCCTTCCGATGACCTAATGATTCTGAGCTTGGCAAAGGTCTTATCTCCCAGCTCGCCCAGGCCCA
GTGTTCCAGGAATGTGACCTTTGCTGCAGCAGCCGCTGGAGGGGGCAGAGGGGATGGGCTGGAGGTTGAGCAAACAG
AGCAGCAGAAAAGGCAGTTCCTCTTCTCCAGTGCCCTCCTTCCCTGTCTCTGCCTCTCCCTCCCTTCCTCAGGCATC
AGAGCGGAGACTTCAGGGAGACCAGAGCCCAGCTTGCCAGGCACTGAGCTAGAAGCCCTGCCATGGCACCCCTGAGA
CCCCTTCTCATACTGGCCCTGCTGGCATGGGTTGCTCTGG

<210> SEQ ID NO 1040

<211> Length : 168

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1040

>T39971_node_18

GTGCCAGGGGCTGTGGGCCAGGGTAGAAAGCATCTAGGGAGGGTTTGAGAGCTATTGCTCCCAGGGACAGGGTGGAC
AGGGAAGCTGGACCCAGGGCCCTGCAGGACCTGGTGGGAGCTCTGTGAGCACAGGGCAGCCCCAAGACTCCAGGTCC
TGGGCAGTGAACCT

<210> SEQ ID NO 1041

<211> Length : 157

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1041
>T39971_node_21
GGTAGTCAGTACTGGCGCTTTGAGGATGGTGTCCTGGACCCTGATTACCCCCGAAATATCTCTGACGGCTTCGATGG
CATCCCGGACAACGTGGATGCAGCCTTGGCCCTCCCTGCCCATAGCTACAGTGGCCGGGAGCGGGTCTACTTCTTCA
AGG

<210> SEQ ID NO 1042

<211> Length : 198

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1042
>T39971_node_22
GTACTCAGGGGGTGGTGGGAGACTGAGCAGGCAGTGGAGCAGTCTTGGATTCCTTTCACATTTCACTGGGGACAGGC
CTCAGCATGTGCCCACCCCTGACCCCCACCTCATGCTGGGAGATCCTAACTTCAACAGCCTCTGGGATCTCCAGTCT
TGCCCTGGCCCAGCCCTCCTAATGCCCACCACCCCGCTCCTCAG

<210> SEQ ID NO 1043

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1043
>T39971_node_23

GGAAACAGTACTGGGAGTACCAGTTCCAGCACCAGCCCAGTCAGGAGGAGTGTGAAGGCAGCTCCCTGTCGGCTGTG
TTTGAACACTTTGCCATGATGCAGCGGGACAGCTGGGAGGACATCTTCGAGCTTCTCTTCTGGGGCAGAACCTCTG

<210> SEQ ID NO 1044

<211> Length : 140

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1044
>T39971_node_31
GCCATCCCGCGCCACGTGGCTGTCCTTGTTCTCCAGTGAGGAGAGCAACTTGGGAGCCAACAACTATGATGACTACA
GGATGGACTGGCTTGTGCCTGCCACCTGTGAACCCATCCAGAGTGTCTTCTTCTTCTCTGGAG

<210> SEQ ID NO 1045

<211> Length : 127

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1045
>T39971_node_33
ACAAGTACTACCGAGTCAATCTTCGCACACGGCGAGTGGACACTGTGGACCCTCCCTACCCACGCTCCATCGCTCAG
TACTGGCTGGGCTGCCCAGCTCCTGGCCATCTGTAGGAGTCAGAGCCCAC

<210> SEQ ID NO 1046

<211> Length : 223

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1046
>T39971_node_7
TGACTCGCGGGGATGTGTTCACTATGCCGGAGGATGAGTACACGGTCTATGACGATGGCGAGGAGAAAAACAATGCC
ACTGTCCATGAACAGGTGGGGGGCCCCTCCCTGACCTCTGACCTCCAGGCCCAGTCCAAAGGGAATCCTGAGCAGAC
ACCTGTTCTGAAACCTGAGGAAGAGGCCCCTGCGCCTGAGGTGGGCGCCTCTAAGCCTGAGGGGATAGA

<210> SEQ ID NO 1047
<211> Length : 9
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1047
>T39971_node_1
CTGACCAAG

<210> SEQ ID NO 1048
<211> Length : 44
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1048
>T39971_node_10
GGAGACCTCAGCCCCCAGCAGAGGAGGAGCTGTGCAGTGGGAAG

<210> SEQ ID NO 1049
<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1049
>T39971_node_11
CCCTTCGACGCCTTCACCGACCTCAAGAACGGTTCCCT

<210> SEQ ID NO 1050

<211> Length : 14

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1050
>T39971_node_12
CTTTGCCTTCCGAG

<210> SEQ ID NO 1051

<211> Length : 32

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1051
>T39971_node_15
GGCAGTACTGCTATGAACTGGACGAAAAGGCA

<210> SEQ ID NO 1052

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1052
>T39971_node_16
GTGAGGCCTGGGTACCCCAAGCTC

<210> SEQ ID NO 1053

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1053
>T39971_node_17
ATCCGAGATGTCTGGGGCATCGAGGGCCCCATCGATGCCGCCTTCACCCGCATCAACTGTCAGGGGAAGACCTACCT
CTTCAAG

<210> SEQ ID NO 1054

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1054
>T39971_node_26
CTGGTACCAGACAGCCCCAGTTCATTAGCCGGGACTGGCACG

<210> SEQ ID NO 1055

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1055
>T39971_node_27
GTGTGCCAGGGCAAGTGGACGCAGCCATGGCTGGCCGCATCTACATCTCAG

<210> SEQ ID NO 1056

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1056
>T39971_node_28
GCATGGCAC

<210> SEQ ID NO 1057

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1057
>T39971_node_29
CCCGCCCCTCCTTGGCCAAGAAACAAAGGTTTAGGCATCGCAACCGCAAAGGCTACCGTTCACAACGAGGCCACAGC
CGTGGCCGCAACCAGAAC

<210> SEQ ID NO 1058

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1058
>T39971_node_3
AGTCATGCAAGGGCCGCTGCACTGAGGGCTTCAACGTGGACA

<210> SEQ ID NO 1059

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1059
>T39971_node_30
TCCCGCCG

<210> SEQ ID NO 1060

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1060
>T39971_node_34
ATGGCCG

<210> SEQ ID NO 1061

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1061
>T39971_node_35
GGCCCTCTGTAGCTCCC

<210> SEQ ID NO 1062

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1062
>T39971_node_36
TCCTCCCATCTCCTTCCCCCAGCCCAATAAAGGTCCCTTAGCCCCGAAAAAAAAAGCKATAAT

<210> SEQ ID NO 1063

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1063
>T39971_node_4
AGAAGTGCCAGTGTGACGAG

<210> SEQ ID NO 1064

<211> Length : 58

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1064
>T39971_node_5
CTCTGCTCTTACTACCAGAGCTGCTGCACAGACTATACGGCTGAGTGCAAGCCCCAAG

<210> SEQ ID NO 1065

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1065
>T39971_node_8
CTCAAG

<210> SEQ ID NO 1066

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1066
>T39971_node_9

GCCTGAGACCCTTCATCCAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1067

<211> Length : 327

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1067

>Z21368_PEA_1_node_0

AGGTTACTTGACTGGGAGTTCTCAGACCTCCAGTTTCAGCCCTGCCCTCAGCCTCCAATCCGTAAGAGACACCCAGC

CCCAGCAATTGGATTGGGCAGCCCGTCTTGACACACCACTGTGCTGAGTGCTTGAGGACGTGTTTCAACAGATGGTT

GGGGTTAGTGTGTGTCATCACATTCGAGTGGGGATTAAGAGAAGGAAGGCTGCCTTGCTGGAGCTGTGTGGTCTTCT

CCAAGTGAGAGTCGCAGGCAATAGAACTACTTTGCTTTTGGAGGAAAAGGAGGAATTCATTTTCAGCAGACACAAGA

AAAGCAGTTTTTTTTTCAG

<210> SEQ ID NO 1068

<211> Length : 177

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1068

>Z21368_PEA_1_node_15

AACTCCAGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATGAAGTATTCTTGCTG

TGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGTTCAGAG

GACGGATACAGCAGGAACGAAAA

<210> SEQ ID NO 1069

<211> Length : 240

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1069

>Z21368_PEA_1_node_19

GGTCCCTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATCAATGCCTTTGTGACT
ACACCCATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAATGTCTACACCAACAA
CGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATCTTAACAACACTGGCT
ACAGAACAG

<210> SEQ ID NO 1070

<211> Length : 300

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1070

>Z21368_PEA_1_node_2

TCTTCATCTTGCGAGCACTTGGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGG
CAACTAGGAAACCCCAGGCGCAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGC
GCTCCTCGGAGGTCAGGGCAGATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTA
TCTCTGCTCCTCTGCCTCCTCCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTT

<210> SEQ ID NO 1071

<211> Length : 152

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1071

>Z21368_PEA_1_node_21

CCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGGTGGCGAGAATGGCTTGGATTAATC

AAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAAGCATGGATTTGATTATGCAAAG

<210> SEQ ID NO 1072

<211> Length : 176

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1072

>Z21368_PEA_1_node_33

CTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCATTTACACCGCCGACCATGGTTACCATAT

TGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTC

CAAGTGTAGAACCAGGATCAAT

<210> SEQ ID NO 1073

<211> Length : 129

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1073

>Z21368_PEA_1_node_36

AGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCTCCTGATG

TGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAG

<210> SEQ ID NO 1074

<211> Length : 279

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1074
>Z21368_PEA_1_node_37
GTGTGTCATTGTTCCTCCTCTCAGCCAGCCCCAAATACACTGAGCTCCAGCTGGTGCCCAGAGCCAGCCAGCAGCTG
AAGACATGGAGGCAGAATATGCCTTGCCCACAAGGATCACCCCAAGCTGAGCATTTCTCAGCTGCTTGTGAATAGCA
TATTGATGGAGATGCACTCATGGTCTGTGGGAAGTGAGAGGTGTTTCTTTAAATAAGCTGTTAGCACAGATCCATTT
GGAAAAACGTCCAGATGCCAAAAGTAAATATTATCATTTTGCTTTCAG

<210> SEQ ID NO 1075

<211> Length : 853

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1075
>Z21368_PEA_1_node_39
GTAATTATTGGTTCCTGGGGTGCTTCTGGGAACCAGTCCTAGTGGGCAGCTTTCCCTGCTGAGTATTTTTTTTCTCC
TTATTTTTGTTTACTAAGCATGCAGATTTCGTAAACCTAGTCACAAGATTGAATGGTTTGCTGCTTATTCTGTAGTG
GTCAATAGAGTAATAATTGCTGGATCAGAATTGTAAAGAATAACCCTCAAGTTGGTTAATTGGTACAAAAACACAGT
TAGATAGAAGTTATAGAATTTGATAGTATAGTTGGGACATTATCGTTAACAATAATTTATGTATATCTTAAAATAGC
TAGAAGTGAAGAATTGCAAAGTTCCCAACACAAGGAAAAGATAAATGAGATGATGAATATCCCAATTATCTTGATTT
GATCATTACACATTGTAGACTGGTATCCATATATCACACGTACCCCCAAAATATGTATAATTGTGATATATCAATTT
TTAAAATACCAAAAAAGCAAGAGAATGACGACTCCACATCCCCCAAAAAGAATAAATTCTCATAAGCTTGGACCAAA
GCCTTTATCATGGGTGTAGATTTACTGTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTC
TACCATCCTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCTTCTTCCTACTTCCTACTTCTTT
TCCTTGTGTGCCTTTCCTAGTTTTAAATAGATAAATGTATGCCATTGTAATTATTTCCATTGTCACTTCTGGGTTTC

CCCTTTTGGTTCATTAATACCCATTGCCTTGTTTTTCTCTGTACATAAATTAGGAGAGAGAAAATATTTGTATAATT
TTTTTA

<210> SEQ ID NO 1076

<211> Length : 162

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1076
>Z21368_PEA_1_node_4
GGATTCTTCACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTAT
TTAGTCTACAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGA
ATGGAGAG

<210> SEQ ID NO 1077

<211> Length : 130

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1077
>Z21368_PEA_1_node_41
CAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCA
AAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAG

<210> SEQ ID NO 1078

<211> Length : 217

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1078
>Z21368_PEA_1_node_43
AAGTGGCAATGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGT
CCGGCAGAGCACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTT
ACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCAA

<210> SEQ ID NO 1079

<211> Length : 256

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1079
>Z21368_PEA_1_node_45
AGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACATA
AATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGGG
GCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCAC
CTACCACTGTCCGAGTGACACACAA

<210> SEQ ID NO 1080

<211> Length : 176

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1080
>Z21368_PEA_1_node_53
GGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGAGGAAGG
AGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAACCACTGG
CAGACAGCCCCGTTCTGGAACC

<210> SEQ ID NO 1081

<211> Length : 143

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1081

>Z21368_PEA_1_node_56

TGGGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAAT

TTTCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAG

<210> SEQ ID NO 1082

<211> Length : 124

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1082

>Z21368_PEA_1_node_58

CTCACAAATACAGTGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTG

TCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATCTTGATGTTG

<210> SEQ ID NO 1083

<211> Length : 588

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1083

>Z21368_PEA_1_node_66

AGGACAGTTATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAG

CTACACAGTGTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTT

GAAGGATTTAGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTC

AAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGAC

CCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAAC

ATTAAGTCCAGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAA
AAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTA
GAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAG

<210> SEQ ID NO 1084

<211> Length : 585

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1084

>Z21368_PEA_1_node_67

AACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATT
TCAGTTCATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAA
GGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGA
TGAAACTGTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGC
CACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACG
GAGACTCATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGA
TTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGACATAAGTATATACATGT
TATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAA

<210> SEQ ID NO 1085

<211> Length : 1,188

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1085

>Z21368_PEA_1_node_69

TTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAA
ATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGAC
AAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATG
TTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAA

TAAATAAGTAAACAAAATGAAGATTGCCTGCTCTCTCTGTGCCTAGCCTCAAAGCGTTCATCATACATCATACCTTT
AAGATTGCTATATTTTGGGTTATTTTCTTGACAGGAGAAAAAGATCTAAAGATCTTTTATTTTCATCTTTTTTGGTT
TTCTTGGCATGACTAAGAAGCTTAAATGTTGATAAAATATGACTAGTTTTGAATTTACACCAAGAACTTCTCAATAA
AAGAAAATCATGAATGCTCCACAATTTCAACATACCACAAGAGAAGTTAATTTCTTAACATTGTGTTCTATGATTAT
TTGTAAGACCTTCACCAAGTTCTGATATCTTTTAAAGACATAGTTCAAAATTGCTTTTGAAAATCTGTATTCTTGAA
AATATCCTTGTTGTGTATTAGGTTTTTAAATACCAGCTAAAGGATTACCTCACTGAGTCATCAGTACCCTCCTATTC
AGCTCCCCAAGATGATGTGTTTTTGCTTACCCTAAGAGAGGTTTTCTTCTTATTTTTAGATAATTCAAGTGCTTAGA
TAAATTATGTTTTCTTTAAGTGTTTATGGTAAACTCTTTTAAAGAAAATTTAATATGTTATAGCTGAATCTTTTTGG
TAACTTTAAATCTTTATCATAGACTCTGTACATATGTTCAAATTAGCTGCTTGCCTGATGTGTGTATCATCGGTGGG
ATGACAGAACAAACATATTTATGATCATGAATAATGTGCTTTGTAAAAAGATTTCAAGTTATTAGGAAGCATACTCT
GTTTTTTAATCATGTATAATATTCCATGATACTTTTATAGAACAATTCTGGCTTCAGGAAAGTCTAGAAGCAATATT
TCTTCAAATAAAAGGTGTTTAAACTTTTTTCTG

<210> SEQ ID NO 1086

<211> Length : 45

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1086

>Z21368_PEA_1_node_11

GACCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTG

<210> SEQ ID NO 1087

<211> Length : 28

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1087

>Z21368_PEA_1_node_12

ATTATTCAACCAGGATACCTAATTCAAG

<210> SEQ ID NO 1088

<211> Length : 15

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1088
>Z21368_PEA_1_node_16
AACATCCGACCCAAC

<210> SEQ ID NO 1089

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1089
>Z21368_PEA_1_node_17
ATTATTCTTGTGCTTACCGATGATCAAGATGTGGAGCTGG

<210> SEQ ID NO 1090

<211> Length : 74

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1090

>Z21368_PEA_1_node_23
GACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTAAGAGAATGTATCCCCATAG

<210> SEQ ID NO 1091

<211> Length : 96

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1091
>Z21368_PEA_1_node_24
GCCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTCTAAACTGTACC
CCAATGCTTCCCAACACAT

<210> SEQ ID NO 1092

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1092
>Z21368_PEA_1_node_30
AACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTATGCAGTACACAG

<210> SEQ ID NO 1093

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1093

>Z21368_PEA_1_node_31
GACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTGATGTCAGTGGAT
GATTCTGTGGAGAGG

<210> SEQ ID NO 1094

<211> Length : 57

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1094
>Z21368_PEA_1_node_38
GTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGG

<210> SEQ ID NO 1095

<211> Length : 97

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1095
>Z21368_PEA_1_node_47
GTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATA
AGGCATACATTGACAAAGAG

<210> SEQ ID NO 1096

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1096
>Z21368_PEA_1_node_49
ATTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATG
TAGCTGCAGTAAACAAAG

<210> SEQ ID NO 1097

<211> Length : 66

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1097
>Z21368_PEA_1_node_51
CTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCCATTCAA

<210> SEQ ID NO 1098

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1098
>Z21368_PEA_1_node_61
GAAATAAAGATGGAGGAAGCTATGACCTACACAG

<210> SEQ ID NO 1099

<211> Length : 53

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1099
>Z21368_PEA_1_node_68
CTTGACACCCCTGGTAAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAG

<210> SEQ ID NO 1100

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1100
>Z21368_PEA_1_node_7
GTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAAT
TGCAGAAAATCTTCAAAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1101

<211> Length : 148

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1101
>HSKITCR_node_0
GGCTCGGCTTTGCCGCGCTCGCTGCACTTGGGCGAGAGCTGGAACGTGGACCAGAGCTCGGATCCCATCGCAGCTAC
CGCGATGAGAGGCGCTCGCGGCGCCTGGGATTTTCTCTGCGTTCTGCTCCTACTGCTTCGCGTCCAGACAG

<210> SEQ ID NO 1102

<211> Length : 190

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1102
>HSKITCR_node_11
ATAAAGGATTCATTAATATCTTCCCCATGATAAACACTACAGTATTTGTAAACGATGGAGAAAATGTAGATTTGATT
GTTGAATATGAAGCATTCCCCAAACCTGAACACCAGCAGTGGATCTATATGAACAGAACCTTCACTGATAAATGGGA
AGATTATCCCAAGTCTGAGAATGAAAGTAATATCAG

<210> SEQ ID NO 1103

<211> Length : 194

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1103
>HSKITCR_node_17
ATGCTCTGCTTCTGTACTGCCAGTGGATGTGCAGACACTAAACTCATCTGGGCCACCGTTTGGAAAGCTAGTGGTTC
AGAGTTCTATAGATTCTAGTGCATTCAAGCACAATGGCACGGTTGAATGTAAGGCTTACAACGATGTGGGCAAGACT
TCTGCCTATTTTAACTTTGCATTTAAAGGTAACAACAAAG

<210> SEQ ID NO 1104

<211> Length : 270

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1104
>HSKITCR_node_2
GCTCTTCTCAACCATCTGTGAGTCCAGGGGAACCGTCTCCACCATCCATCCATCCAGGAAAATCAGACTTAATAGTC
CGCGTGGGCGACGAGATTAGGCTGTTATGCACTGATCCGGGCTTTGTCAAATGGACTTTTGAGATCCTGGATGAAAC
GAATGAGAATAAGCAGAATGAATGGATCACGGAAAAGGCAGAAGCCACCAACACCGGCAAATACACGTGCACCAACA
AACACGGCTTAAGCAATTCCATTTATGTGTTTGTTAGAG

<210> SEQ ID NO 1105
<211> Length : 127
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1105
>HSKITCR_node_21
AAACCCATGTATGAAGTACAGTGGAAGGTTGTTGAGGAGATAAATGGAAACAATTATGTTTACATAGACCCAACACA
ACTTCCTTATGATCACAAATGGGAGTTTCCCAGAAACAGGCTGAGTTTTG

<210> SEQ ID NO 1106
<211> Length : 151
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1106
>HSKITCR_node_27
GGCCCACCCTGGTCATTACAGAATATTGTTGCTATGGTGATCTTTTGAATTTTTTGAGAAGAAAACGTGATTCATTT
ATTTGTTCAAAGCAGGAAGATCATGCAGAAGCTGCACTTTATAAGAATCTTCTGCATTCAAAGGAGTCTTCCTG

<210> SEQ ID NO 1107

<211> Length : 125

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1107
>HSKITCR_node_3
GTAAATGCTTGGCTTTCTGCAGTGCTGTGCTTTCAAGAATTTAATATCCTGCTCTTAATTTTGGATGACATATGGAT
GACTGAGCCATAGATAAAATATTTCTGGCTGGGTCTAGAAGGCCTAAA

<210> SEQ ID NO 1108

<211> Length : 128

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1108
>HSKITCR_node_31
GCTCATACATAGAAAGAGATGTGACTCCCGCCATCATGGAGGATGACGAGTTGGCCCTAGACTTAGAAGACTTGCTG
AGCTTTTCTTACCAGGTGGCAAAGGGCATGGCTTTCCTCGCCTCCAAGAAT

<210> SEQ ID NO 1109

<211> Length : 123

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1109
>HSKITCR_node_33
TGTATTCACAGAGACTTGGCAGCCAGAAATATCCTCCTTACTCATGGTCGGATCACAAAGATTTGTGATTTTGGTCT
AGCCAGAGACATCAAGAATGATTCTAATTATGTGGTTAAAGGAAAC

<210> SEQ ID NO 1110

<211> Length : 1,321

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1110

>HSKITCR_node_34

GTGAGTACCCATTCTCTGCTTGACAGTCCTGCAAAGGATTTTTAGTTTCAACTTTCGATAAAAATTGTTTCCTGTGA
CTTTCATAATGTAAATCCTGTCTAGGGATATCACACATTTTAGCAGTCAAATTAAGTATACTTCAGCAAAATTTGCA
TGGTATGCTGAACATTACTACAACTAACATTCAATAATAGAAGTCCTAATTCTAATTGTGTAATTTTGGGGCATGTG
AAGGAAACAGAAATAGCCTTAATTTTCATTATAGCCTGAGAATAGCAATGAACTTGATTTTGCTCAAGTGTAACAAA
TGTAGGTCATTGAAGGTCACAGCAGGAGAAATTTTGGGGGGATTGGCATGCCGTGTGAAAAATATTAAAATCTAAGA
TCATATTCAGAGTTAGCCATATAGAATGTTGGATCCTAGAATACACGGAGAGCTATTAAATAGGTTCATAAGTAATA
ATGGGTTTGTAGAACATAGCAAATGTTTTTCAAAGTGCTGTATTTCCTTGCCAATTTTTATTGAATCCTCAGAACAA
ACCCTTGTCCTATAATTGCAGCACCTTCCTATTTTAAAAGTTGGAGAAACTAAGGCATGGCATATTTTAGTGCTTTT
TATCTGAGGTGACACAGGACAATAGAGGCTGAAACTAGAGCCTGTGATTTCCAGAGTGGCATCCCATCGCCATGGAA
AACCATATTGCCCTTTATAATATTCTCTGCCTGTCTTCCTGGGCCCATTTCTGCAGTCCAGATTTATTCCTGGAGGT
GGATTTTTGTGCCCCTTATGATAATGTAATGTTTTCCCCCCATCACTGTGGCGTCTCTATTGGTAGCCTGTACTCGA
TTTGGTTTTGTAAGAGTGTTTCTGTGTTTATCTCCTGGCCAAAGCATGAATTTCTGAAAAGGCTGCCCTGAAGCTAG
ATATTATCATCAGGAGTGATAATAAAAACAACTTCTCTTCATCCTGGTGCAATTTTGGAAGGAAGAATGATAAGTCG
TTAATTATCCAGTCCTCTGAATATTGACCAAAGGGAGGAGGAAGAAGGTGGCCGCCTCCAGGCCCAGTCTTGATGAA
TTCTGCAATCTGTCTCCATTATTTCATGCCTGTGTAATTTTGTCAAGAGTTGTCCAATACCTCTGAATCTTTGTTTT
TCTCCCCTTGGGAATAATAGTGATTCCTGACTCAGAAAGTTATGAGTTAATATATGTAAAGTGCTTATGACAGTGCC
AGACATATTGTTACTATTCTTTATAGATGTCTTATTTACGATCTAAATAATCGTGACGGTGGTTACATACTGTCTTG
TTATTGGTTAGT

<210> SEQ ID NO 1111

<211> Length : 194

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1111

>HSKITCR_node_36

AATGCACTTGTGAGCCATGTATTTCAGAGGTGATTGGGATCATCTGAGTTCATATAGGTAAAAGGTTTTTGTGAGAT
GGTACTCAAGTTATCACTCCACATTTCAGCAACAGCAGCATCTATAAGAATATCTTCTGTTCAATTTTGTTGAGCTT
CTGAATTAACATTATTGACTCTGTTGTGCTTCTATTACAG

<210> SEQ ID NO 1112

<211> Length : 1,494

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1112

>HSKITCR_node_44

GTAGACCATTCTGTGCGGATCAATTCTGTCGGCAGCACCGCTTCCTCCTCCCAGCCTCTGCTTGTGCACGACGATGT
CTGAGCAGAATCAGTGTTTGGGTCACCCCTCCAGGAATGATCTCTTCTTTTGGCTTCCATGATGGTTATTTTCTTTT
CTTTCAACTTGCATCCAACTCCAGGATAGTGGGCACCCCACTGCAATCCTGTCTTTCTGAGCACACTTTAGTGGCCG
ATGATTTTTGTCATCAGCCACCATCCTATTGCAAAGGTTCCAACTGTATATATTCCCAATAGCAACGTAGCTTCTAC
CATGAACAGAAAACATTCTGATTTGGAAAAAGAGAGGGAGGTATGGACTGGGGGCCAGAGTCCTTTCCAAGGCTTCT
CCAATTCTGCCCAAAAATATGGTTGATAGTTTACCTGAATAAATGGTAGTAATCACAGTTGGCCTTCAGAACCATCC
ATAGTAGTATGATGATACAAGATTAGAAGCTGAAAACCTAAGTCCTTTATGTGGAAAACAGAACATCATTAGAACAA
AGGACAGAGTATGAACACCTGGGCTTAAGAAATCTAGTATTTCATGCTGGGAATGAGACATAGGCCATGAAAAAAAT
GATCCCCAAGTGTGAACAAAAGATGCTCTTCTGTGGACCACTGCATGAGCTTTTATACTACCGACCTGGTTTTTAAA
TAGAGTTTGCTATTAGAGCATTGAATTGGAGAGAAGGCCTCCCTAGCCAGCACTTGTATATACGCATCTATAAATTG
TCCGTGTTCATACATTTGAGGGGAAAACACCATAAGGTTTCGTTTCTGTATACAACCCTGGCATTATGTCCACTGTG
TATAGAAGTAGATTAAGAGCCATATAAGTTTGAAGGAAACAGTTAATACCATTTTTTAAGGAAACAATATAACCACA
AAGCACAGTTTGAACAAAATCTCCTCTTTTAGCTGATGAACTTATTCTGTAGATTCTGTGGAACAAGCCTATCAGCT
TCAGAATGGCATTGTACTCAATGGATTTGATGCTGTTTGACAAAGTTACTGATTCACTGCATGGCTCCCACAGGAGT
GGGAAAACACTGCCATCTTAGTTTGGATTCTTATGTAGCAGGAAATAAAGTATAGGTTTAGCCTCCTTCGCAGGCAT
GTCCTGGACACCGGGCCAGTATCTATATATGTGTATGTACGTTTGTATGTGTGTAGACAAATATTTGGAGGGGTATT
TTTGCCCTGAGTCCAAGAGGGTCCTTTAGTACCTGAAAAGTAACTTGGCTTTCATTATTAGTACTGCTCTTGTTTCT
TTTCACATAGCTGTCTAGAGTAGCTTACCAGAAGCTTCCATAGTGGTGCAGAGGAAGTGGAAGGCATCAGTCCCTAT
GTATTTGCAGTTCACCTGCACTTAAGGCACTCTGTTATTTAGACTCATCTTACTGTACCTGTTCCTTAGACCTTCCA
TAATGCTACTGTCTCACTGAAACATTTAAAT

<210> SEQ ID NO 1113

<211> Length : 402

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1113

>HSKITCR_node_46

TTTACCCTTTAGACTGTAGCCTGGATATTATTCTTGTAGTTTACCTCTTTAAAAACAAAACAAAACAAAACAAAAAA

CTCCCCTTCCTCACTGCCCAATATAAAAGGCAAATGTGTACATGGCAGAGTTTGTGTGTTGTCTTGAAAGATTCAGG

TATGTTGCCTTTATGGTTTCCCCCTTCTACATTTCTTAGACTACATTTAGAGAACTGTGGCCGTTATCTGGAAGTAA

CCATTTGCACTGGAGTTCTATGCTCTCGCACCTTTCCAAAGTTAACAGATTTTGGGGTTGTGTTGTCACCCAAGAGA

TTGTTGTTTGCCATACTTTGTCTGAAAAATTCCTTTGTGTTTCTATTGACTTCAATGATAGTAAGAAAAGTGGTTGT

TAGTTATAGATGTCTAG

<210> SEQ ID NO 1114

<211> Length : 282

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1114

>HSKITCR_node_5

ATCCTGCCAAGCTTTTCCTTGTTGACCGCTCCTTGTATGGGAAAGAAGACAACGACACGCTGGTCCGCTGTCCTCTC

ACAGACCCAGAAGTGACCAATTATTCCCTCAAGGGGTGCCAGGGGAAGCCTCTTCCCAAGGACTTGAGGTTTATTCC

TGACCCCAAGGCGGGCATCATGATCAAAAGTGTGAAACGCGCCTACCATCGGCTCTGTCTGCATTGTTCTGTGGACC

AGGAGGGCAAGTCAGTGCTGTCGGAAAAATTCATCCTGAAAGTGAGGCCAG

<210> SEQ ID NO 1115

<211> Length : 195

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1115

>HSKITCR_node_50

CAATAATGTCTTTTGAATATTCCCAAGCCCATGAGTCCTTGAAAATATTTTTTATATATACAGTAACTTTATGTGTA

AATACATAAGCGGCGTAAGTTTAAAGGATGTTGGTGTTCCACGTGTTTTATTCCTGTATGTTGTCCAATTGTTGACA

GTTCTGAAGAATTCTAATAAAATGTACATATATAAATCAAG

<210> SEQ ID NO 1116

<211> Length : 137

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1116

>HSKITCR_node_7

CCTTCAAAGCTGTGCCTGTTGTGTCTGTGTCCAAAGCAAGCTATCTTCTTAGGGAAGGGGAAGAATTCACAGTGACG

TGCACAATAAAAGATGTGTCTAGTTCTGTGTACTCAACGTGGAAAAGAGAAAACAGTCAG

<210> SEQ ID NO 1117

<211> Length : 169

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1117

>HSKITCR_node_9

ACTAAACTACAGGAGAAATATAATAGCTGGCATCACGGTGACTTCAATTATGAACGTCAGGCAACGTTGACTATCAG
TTCAGCGAGAGTTAATGATTCTGGAGTGTTCATGTGTTATGCCAATAATACTTTTGGATCAGCAAATGTCACAACAA
CCTTGGAAGTAGTAG

<210> SEQ ID NO 1118

<211> Length : 116

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1118
>HSKITCR_node_13
ATACGTAAGTGAACTTCATCTAACGAGATTAAAAGGCACCGAAGGAGGCACTTACACATTCCTAGTGTCCAATTCTG
ACGTCAATGCTGCCATAGCATTTAATGTTTATGTGAATA

<210> SEQ ID NO 1119

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1119
>HSKITCR_node_15
CAAAACCAGAAATCCTGACTTACGACAGGCTCGTGAATGGCATGCTCCAATGTGTGGCAGCAGGATTCCCAGAGCCC
ACAATAGATTGGTATTTTTGTCCAGGAACTGAGCAGAG

<210> SEQ ID NO 1120

<211> Length : 107

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1120
>HSKITCR_node_19
AGCAAATCCATCCCCACACCCTGTTCACTCCTTTGCTGATTGGTTTCGTAATCGTAGCTGGCATGATGTGCATTATT
GTGATGATTCTGACCTACAAATATTTACAG

<210> SEQ ID NO 1121

<211> Length : 105

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1121
>HSKITCR_node_23
GGAAAACCCTGGGTGCTGGAGCTTTCGGGAAGGTTGTTGAGGCAACTGCTTATGGCTTAATTAAGTCAGATGCGGCC
ATGACTGTCGCTGTAAAGATGCTCAAGC

<210> SEQ ID NO 1122

<211> Length : 111

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1122
>HSKITCR_node_25
CGAGTGCCCATTTGACAGAACGGGAAGCCCTCATGTCTGAACTCAAAGTCCTGAGTTACCTTGGTAATCACATGAAT
ATTGTGAATCTACTTGGAGCCTGCACCATTGGAG

<210> SEQ ID NO 1123

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1123
>HSKITCR_node_29
CAGCGATAGTACTAATGAGTACATGGACATGAAACCTGGAGTTTCTTATGTTGTCCCAACCAAGGCCGACAAAAGGA
GATCTGTGAGAATAG

<210> SEQ ID NO 1124

<211> Length : 112

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1124
>HSKITCR_node_37
GCTCGACTACCTGTGAAGTGGATGGCACCTGAAAGCATTTTCAACTGTGTATACACGTTTGAAAGTGACGTCTGGTC
CTATGGGATTTTTCTTTGGGAGCTGTTCTCTTTAG

<210> SEQ ID NO 1125

<211> Length : 100

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1125
>HSKITCR_node_39
GAAGCAGCCCCTATCCTGGAATGCCGGTCGATTCTAAGTTCTACAAGATGATCAAGGAAGGCTTCCGGATGCTCAGC
CCTGAACACGCACCTGCTGAAAT

<210> SEQ ID NO 1126

<211> Length : 106

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1126

>HSKITCR_node_41

GTATGACATAATGAAGACTTGCTGGGATGCAGATCCCCTAAAAAGACCAACATTCAAGCAAATTGTTCAGCTAATTG

AGAAGCAGATTTCAGAGAGCACCAATCAT

<210> SEQ ID NO 1127

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1127

>HSKITCR_node_43

ATTTACTCCAACTTAGCAAACTGCAGCCCCAACCGACAGAAGCCCGTG

<210> SEQ ID NO 1128

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1128
>HSKITCR_node_47
GTACTTCAGGGGCACTTCATTGAGAGTTTTGTCTTG

<210> SEQ ID NO 1129

<211> Length : 113

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1129
>HSKITCR_node_48
GATATTCTTGAAAGTTTATATTTTTATAATTTTTTCTTACATCAGATGTTTCTTTGCAGTGGCTTAATGTTTGAAAT
TATTTTGTGGCTTTTTTTGTAAATATTGAAATGTAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1130

<211> Length : 760

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1130
>HUMGRP5E_node_0
CCAAAATCTATGGGCTGGGACAGCAAAGATGTGGCCTACGAAGAGAAAGGTCTGGAGAATCAGAAGGCCTTCAAATG
GTGGTTCCAAATCCCTCCAGCAAAGCCCATCCATCTTTAGAGCTCACCCGTCTCCAGCTACACCCCCCACCCCTCCC
GGCCCAGATCAGGCAGCGGGGTCGCCCTCTCCAGGACTCTCAAGGCAGCTAAGGCTGGAGGCGCCGGCGAGCCTGGA
GAGGGAGGAGTTCACTAAATTGTGTTGGATGGAAGGCGTCGAGGACCGGAGGAATTAATCCGATGTGGGGAAGGCGG
ACGGGGCTACGAGGAAAAAAGAGGGGGCAATGTACACTCAGCCTTTTCATCACTCGGCGGGGAGATGGATGGTTTTC
CGGACCGGGCGTCCCAGCGCCCCGGTTAGCTATAGGGAGACGTCAGAGCGCTCTGGTCCGCGATAGAAGAGCCCCCC
AGCCCCCCCGCCCGGGCTTCCATATAAAGTAGGGGCCCTAGTGGAGGCCGCAGCAGTAGCACCAGCGGCTGCGGCGG

CGGAGCTCCTCCGAGGTCCGGGTCACCAGTCTCTGCTCTTCCCAGCCTCTCCGGCGCGCTCCAAGGGCTTCCCGTCG
GGACCATGCGCGGCAGTGAGCTCCCGCTGGTCCTGCTGGCGCTGGTCCTCTGCCTGGCGCCCCGGGGGCGAGCGGTC
CCGCTGCCTGCGGGCGGAGGGACCGTGCTGACCAAGATGTACCCGCGCGGCAACCACTGGGCGGTGG

<210> SEQ ID NO 1131

<211> Length : 224

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1131
>HUMGRP5E_node_2
GGCACTTAATGGGGAAAAAGAGCACAGGGGAGTCTTCTTCTGTTTCTGAGAGAGGGAGCCTGAAGCAGCAGCTGAGA
GAGTACATCAGGTGGGAAGAAGCTGCAAGGAATTTGCTGGGTCTCATAGAAGCAAAGGAGAACAGAAACCACCAGCC
ACCTCAACCCAAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAGGATAGCAGCAACTTCAAAGAT

<210> SEQ ID NO 1132

<211> Length : 359

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1132
>HUMGRP5E_node_8
GTTCTCAACGTGAAGGAAGGAACCCCCAGCTGAACCAGCAATGATAATGATGGCCTCTCTCAAAAGAGAAAAACAAA
ACCCCTAAGAGACTGCGTTCTGCAAGCATCAGTTCTACGGATCATCAACAAGATTTCCTTGTGCAAAATATTTGACT
ATTCTGTATCTTTCATCCTTGACTAAATTCGTGATTTTCAAGCAGCATCTTCTGGTTTAAACTTGTTTGCTGTGAAC
AATTGTCGAAAAGAGTCTTCCAATTAATGCTTTTTTATATCTAGGCTACCTGTTGGTTAGATTCAAGGCCCCGAGCT
GTTACCATTCACAATAAAAGCTTAAACACATTGTCCAAAGGGCAGGCTGTT

<210> SEQ ID NO 1133

<211> Length : 19

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1133
>HUMGRP5E_node_3
GTAGGTTCAAAAGGCAAAG

<210> SEQ ID NO 1134

<211> Length : 14

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1134
>HUMGRP5E_node_7
ACTCTCTGCTCCAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1135

<211> Length : 178

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1135
>D56406_PEA_1_node_0

TTCACTCACTTTCAAAGCCAGCTGAAGGAAAGAGGAAGTGCTAGAGAGAGCCCCCTTCAGTGTGCTTCTGACTTTTA
CGGACTTGGCTTGTTAGAAGGCTGAAAGATGATGGCAGGAATGAAAATCCAGCTTGTATGCATGCTACTCCTGGCTT
TCAGCTCCTGGAGTCTGTGCTCAG

<210> SEQ ID NO 1136

<211> Length : 780

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1136
>D56406_PEA_1_node_13
TTAATCCAGGAAGATATTCTTGATACTGGAAATGACAAAAATGGAAAGGAAGAAGTCATAAAGAGAAAAATTCCTTA
TATTCTGAAACGGCAGCTGTATGAGAATAAACCCAGAAGACCCTACATACTCAAAAGAGATTCTTACTATTACTGAG
AGAATAAATCATTTATTTACATGTGATTGTGATTCATCATCCCTTAATTAAATATCAAATTATATTTGTGTGAAAAT
GTGACAAACACACTTATCTGTCTCTTCTACAATTGTGGTTTATTGAATGTGATTTTTCTGCACTAATATAAATTAGA
CTAAGTGTTTTCAAATAAATCTAAATCTTCAGCATGATGTGTTGTGTATAATTGGAGTAGATATTAATTAAGTCACC
TGTATAATGTTTTGTAATTTTGCAAAACATATCTTGAGTTGTTTAAACAGTCAAAATGTTTGATATTTTATACCAGC
TTATGAGCTCAAAGTACTACAGCAAAGCCTAGCCTGCATATCATTCACCCAAAACAAAGTAATAGCGCCTCTTTTAT
TATTTTGACTGAATGTTTTATGGAATTGAAAGAAACATACGTTCTTTTCAAGACTTCCTCATGAATCTCTCAATTAT
AGGAAAAGTTATTGTGATAAAATAGGAACAGCTGAAAGATTGATTAATGAACTATTGTTAATTCTTCCTATTTTAAT
GAATGACATTGAACTGAATTTTTTGTCTGTTAAATGAACTTGATAGCTAATAAAAAGACAACTAGCCATCAAAATCA
AAAGTTTCTC

<210> SEQ ID NO 1137

<211> Length : 93

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1137
>D56406_PEA_1_node_11

GCACGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGACGGGCGGATCACGAGGTCAAGAGATGGAGAC
CATCCCGGCTAACACG

<210> SEQ ID NO 1138

<211> Length : 6

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1138
>D56406_PEA_1_node_2
ATTCAG

<210> SEQ ID NO 1139

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1139
>D56406_PEA_1_node_3
AAGAGGAAATGAAAGCATTAGAAGCAGATTTCTTGACCAATATGCATACATCAAAG

<210> SEQ ID NO 1140

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1140

>D56406_PEA_1_node_5
ATTAGTAAAGCACATGTTCCCTCTTGGAAGATGACTCTGCTAAATGTTTGCAGTCTTGTAAATAATTTGAACAGCCC
AGCTGAGGAAACAGGAGAAGTTCATGAAGAGGAGCTTG

<210> SEQ ID NO 1141

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1141
>D56406_PEA_1_node_6
TTGCAAGAAGGAAACTTCCTACTGCTTTAGATGG

<210> SEQ ID NO 1142

<211> Length : 26

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1142
>D56406_PEA_1_node_7
CTTTAGCTTGGAAGCAATGTTGACAA

<210> SEQ ID NO 1143

<211> Length : 8

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1143

>D56406_PEA_1_node_8
TATACCAG

<210> SEQ ID NO 1144

<211> Length : 42

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1144
>D56406_PEA_1_node_9
CTCCACAAAATCTGTCACAGCAGGGCTTTTCAACACTGGGAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1145

<211> Length : 245

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1145
>F05068_PEA_1_node_0
AAGAAAGGGAAGGCAACCGGGCAGCCCAGGCCCCGCCCCGCCGCTCCCCCACCCGTGCGCTTATAAAGCACAGGAAC
CAGAGCTGGCCACTCAGTGGTTTCTTGGTGACACTGGATAGAACAGCTCAAGCCTTGCCACTTCGGGCTTCTCACTG
CAGCTGGGCTTGGACTTCGGAGTTTTGCCATTGCCAGTGGGACGTCTGAGACTTTCTCCTTCAAGTACTTGGCAGAT
CACTCTCTTAGCAG

<210> SEQ ID NO 1146

<211> Length : 161

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1146
>F05068_PEA_1_node_10
CTTCGGGACGTGCACGGTGCAGAAGCTGGCACACCAGATCTACCAGTTCACAGATAAGGACAAGGACAACGTCGCCC
CCAGGAGCAAGATCAGCCCCCAGGGCTACGGCCGCCGGCGCCGGCGCTCCCTGCCCGAGGCCGGCCCGGGTCGGACT
CTGGTGT

<210> SEQ ID NO 1147

<211> Length : 121

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1147
>F05068_PEA_1_node_12
CCATGGTACAAGGAATAGTCGCGCAAGCATCCCGCTGGTGCCTCCCGGGACGAAGGACTTCCCGAGCGGTGTGGGGA
CCGGGCTCTGACAGCCCTGCGGAGACCCTGAGTCCGGGAGGCAC

<210> SEQ ID NO 1148

<211> Length : 631

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1148
>F05068_PEA_1_node_13
CGTCCGGCGGCGAGCTCTGGCTTTGCAAGGGCCCCTCCTTCTGGGGGCTTCGCTTCCTTAGCCTTGCTCAGGTGCAA
GTGCCCCAGGGGGCGGGGTGCAGAAGAATCCGAGTGTTTGCCAGGCTTAAGGAGAGGAGAAACTGAGAAATGAATGC

TGAGACCCCCGGAGCAGGGGTCTGAGCCACAGCCGTGCTCGCCCACAAACTGATTTCTCACGGCGTGTCACCCCACC
AGGGCGCAAGCCTCACTATTACTTGAACTTTCCAAAACCTAAAGAGGAAAAGTGCAATGCGTGTTGTACATACAGAG
GTAACTATCAATATTTAAGTTTGTTGCTGTCAAGATTTTTTTTGTAACTTCAAATATAGAGATATTTTTGTACGTTA
TATATTGTATTAAGGGCATTTTAAAAGCAATTATATTGTCCTCCCCCTATTTTAAGACGTGAATGTCTCAGCGAGGT
GTAAAGTTGTTCGCCGCGTGGAATGTGAGTGTGTTTGTGTGCATGAAAGAGAAAGACTGATTACCTCCTGTGTGGAA
GAAGGAAACACCGAGTCTCTGTATAATCTATTTACATAAAATGGGTGATATGCGAACAGCAAACCAATAAACTGTCT
CAATGCTGAATAAAA

<210> SEQ ID NO 1149

<211> Length : 150

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1149
>F05068_PEA_1_node_4
GTGAGTCCGGGCAGCGCCTTCCCCCTTGCTGGTACCTGGCAGGCAAGGGGAACTGACCGTTGGTCCCGAAGGTCTAG
AAGTGAATGGGAGCAGGGACAGGCCTGGGCGTCACCTGAACGCACGCGAATCGGGTCTGCTTGTGTTTTCCAG

<210> SEQ ID NO 1150

<211> Length : 0

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1150
>F05068_PEA_1_node_8
GTAACTACGCCCTGTGCTGTCCAGGGACGGGAGGGAAGGAAGGTGTGCGGGAGGAGTTCTCTGTCTCCACTCCCCTG
GCCCGGGGGATCGTCGGGGCTGGACCGCAGCTCAGATGGCGCGAGCAGTTTCCAGCTCCCTCTGGCTCTAGAATGGC
TCCCGTTCCCGGTGTTGGGGCCAAAGCTCTGCTTGATGGGGTCTCAAGTTGCCTTTCTTCCCCCTCCCCCCGCCCGC
AG

<210> SEQ ID NO 1151

<211> Length : 76

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1151
>F05068_PEA_1_node_11
CTTCTAAGCCACAAGCACACGGGGCTCCAGCCCCCCCGAGTGGAAGTGCTCCCCACTTTCTTTAGGATTTAGGCGC

<210> SEQ ID NO 1152

<211> Length : 119

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1152
>F05068_PEA_1_node_3
GGTCTGCGCTTCGCAGCCGGGATGAAGCTGGTTTCCGTCGCCCTGATGTACCTGGGTTCGCTCGCCTTCCTAGGCGC
TGACACCGCTCGGTTGGATGTCGCGTCGGAGTTTCGAAAGAA

<210> SEQ ID NO 1153

<211> Length : 59

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1153
>F05068_PEA_1_node_5
GTGGAATAAGTGGGCTCTGAGTCGTGGGAAGAGGGAACTGCGGATGTCCAGCAGCTACC

<210> SEQ ID NO 1154

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1154
>F05068_PEA_1_node_6
CCACCGGGCTCGCTGACGTGAAGGCCGGGCCTGCCCAGAC

<210> SEQ ID NO 1155

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1155
>F05068_PEA_1_node_7
CCTTATTCGGCCCCAGGACATGAAGGGTGCCTCTCGAAGCCCCGAAGACAG

<210> SEQ ID NO 1156

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1156
>F05068_PEA_1_node_9
CAGTCCGGATGCCGCCCGCATCCGAGTCAAGCGCTACCGCCAGAGCATGAACAACTTCCAGGGCCTCCGGAGCTTTG
GCTGCCG

Segment nucleic acid sequences:

<210> SEQ ID NO 1157

<211> Length : 573

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1157

>H14624_node_0

TTATGCTCCCGCGGAGGCCAAGCGGACTCCCTGACAGGACAGAATCTGAACGTGAGAGTGAAGGTCTTGCCTGTCCA
GAAACTCTTGTAGCCAGCACAGGTTTAAACAAGAAGCCAAATTGTTCTGGAGAGATTGCTGGGGGCTTTCTTTGTGC
CTCAAGCTTCTTCAGTGCCCTGAGCACAGGAAACACTCAAGCAGAGAAGCAGAGCCAAACCCAGGATACGGGAGGTC
GAGGCTCTTCCGTAGACCTGCAGCATTGGGGTGGGATGATGTTCATTCTGTGTGTGTTCTGGACCAAGCCCCTCTCC
AGGGACCTATGGGCAGCCCCCTTTAAGCAAGATGCCCGGTGGAGTGGGCATCCACCATCACTTACCCTGGGCTTGGG
TGAATAGATTTTCCGTGCCTTAAATGGGCAGGGAGGGGGTAAACATGGACGGTCCATTGGTACAAATAAAAGCCTTT
GGTGGGTTTTGATCAATTGCAAGGATCGAAGGAGACCTGTGGACCTGAGGTCAACTGGCAGCAGAGAAGAGTCTGGG
TTCGTGAAGGCGCCGCCGCGGTGCCGCGCCACGT

<210> SEQ ID NO 1158

<211> Length : 387

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1158

>H14624_node_16

GTAAGCCTTCCCTCTTGCTTCCCCACTCCCTGCTGGGCTGAGACGCTCCCAGGAGATCCCGCCCCTGCCACGCATCC
CAGTGCATCCCTGCTTGGGGTGCCAGTAGCGGGAAGGGCAGAAGTTCTGCCTGACCTGGTCTGTCATCACAACAAGC
CTGTATCAAATTTGAGGCACCCCTCCCACGCCGCCCAAGTCTCGCGCATTCTCCTCCCGAGTTGTACCAGCTATACT
TAAGGGCAGTTTAAAAATAAAACAAACAAACAAAAACAACAAAACTAAAAAAACGAAGAACTGAACGGCGGTTTAAA
AAAAAATAGATACACGATTATTGTTAAAGATGCTAGCACTGGAGCTGCGCAGAGCGTTGGAAGTGGTGTTTGGTGGA
GG

<210> SEQ ID NO 1159

<211> Length : 249

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1159

>H14624_node_3

ATTTGCATAAAAAAGGCCAAGAAAACTCTGGCTGTGCCCCAGCAACGGCTCATTCTGCTCCCCCGGGTCGGAGCCCC

CCGGAGCTGCGCGCGGGCTTGCAGCGCCTCGCCCGCGCTGTCCTCCCGGTGTCCCGCTTCTCCGCGCCCCAGCCGCC

GGCTGCCAGCTTTTCGGGGCCCCGAGTCGCACCCAGCGAAGAGAGCGGGCCCGGGACAAGCTCGAACTCCGGCCGCC

TCGCCCTTCCCCGGCTCC

<210> SEQ ID NO 1160

<211> Length : 10

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1160

>H14624_node_10

GTGCTGGAGC

<210> SEQ ID NO 1161

<211> Length : 35

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1161

>H14624_node_11

AGGCCGGCGCTTGGATCCCGCTGGTCATGAAGCAG

<210> SEQ ID NO 1162

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1162
>H14624_node_12
TGCCACCCGGACACCAAGAAG

<210> SEQ ID NO 1163

<211> Length : 92

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1163
>H14624_node_13
TTCCTGTGCTCGCTCTTCGCCCCCGTCTGCCTCGATGACCTAGACGAGACCATCCAGCCATGCCACTCGCTCTGCGT
GCAGGTGAAGGACCG

<210> SEQ ID NO 1164

<211> Length : 60

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1164
>H14624_node_14
CTGCGCCCCGGTCATGTCCGCCTTCGGCTTCCCCTGGCCCGACATGCTTGAGTGCGACCG

<210> SEQ ID NO 1165

<211> Length : 71

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1165
>H14624_node_15
TTTCCCCCAGGACAACGACCTTTGCATCCCCCTCGCTAGCAGCGACCACCTCCTGCCAGCCACCGAGGAAG

<210> SEQ ID NO 1166

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1166
>H14624_node_4
GCTCCCTCTGCCCCCTCGGGGTCGCGCGCCCACGATGCTGCAGGGCCCTGGCTCGCTGCTGCTGCTCTTC

<210> SEQ ID NO 1167

<211> Length : 11

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1167
>H14624_node_5
CTCGCCTCGCA

<210> SEQ ID NO 1168

<211> Length : 24

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1168

>H14624_node_6

CTGCTGCCTGGGCTCGGCGCGCGG

<210> SEQ ID NO 1169

<211> Length : 7

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1169

>H14624_node_7

GCTCTTC

<210> SEQ ID NO 1170

<211> Length : 80

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1170

>H14624_node_8

CTCTTTGGCCAGCCCGACTTCTCCTACAAGCGCAGCAATTGCAAGCCCATCCCGGCCAACCTGCAGCTGTGCCACGG
CAT

<210> SEQ ID NO 1171

<211> Length : 55

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1171
>H14624_node_9
CGAATACCAGAACATGCGGCTGCCCAACCTGCTGGGCCACGAGACCATGAAGGAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1172

<211> Length : 213

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1172
>H38804_PEA_1_node_0
GACTGGGTTGACCGATGCTGGGCAGCTGAGCGGACCAATCGGCCCCCTAGACTGAGACGTTGGCGTTTGAAATCAGC
CAATGGCAGGTCTACACTGGAGCTTCCTCTCCGCCTCCTTCGCCTAGCCTGCGAGTGTTCTGAGGGAAGCAAGGAGG
CGGCGGCGGCCGCAGCGAGTGGCGAGTAGTGGAAACGTTGCTTCTGAGGGGAGCCCAAG

<210> SEQ ID NO 1173

<211> Length : 432

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1173
>H38804_PEA_1_node_1
GTAGGGAGGCGAGGCGACGGTGTGCGGGAGCGGGCTCTCCAGGGACTTCCCGGGTCCGCAACTGGCAGGGCCGTTCG
ATTCGCAGGGGATCCCGTTTCGTTTCTGTTGTTTTCCCTTTATTTTTAGGAGTGCCCGGGGCGACGGGACCCCGGGA
GAGGGGAAAGGGAACAGTCTGGGGTCCGGGCATCGCTGTGGGCCGGGCTGGGTTTAGGGGGACGGCGGTGCGGGCTG

GGCCGGTTTGGGCGCGGCGGGGGCCGGATGATGGGGCGAGTCCGGACCTTGGCGGGCGAGTGCTCGGCGCAGGCGCA
AGCGCAGAGTCTCCTCGCGGTCGTCCTCTCGGCCCCTCCCTCTGGGGGGACCCCCAGTGCCAGGCTGTCAGTGCGCA
GCCCCAGCCCGCGGGACCCCTGGGGACTCTGGGCGCCTGTTCTGCAG

<210> SEQ ID NO 1174

<211> Length : 159

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1174
>H38804_PEA_1_node_16
GTATATACCCTCTCAGTGTCTGGAGACCGGCTGATTGTGGGAACAGCAGGCCGCAGAGTGTTGGTGTGGGACTTACG
GAACATGGGTTACGTGCAGCAGCGCAGGGAGTCCAGCCTGAAATACCAGACTCGCTGCATACGAGCGTTTCCAAACA
AGCAG

<210> SEQ ID NO 1175

<211> Length : 139

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1175
>H38804_PEA_1_node_19
GGTTATGTATTAAGCTCTATTGAAGGCCGAGTGGCAGTTGAGTATTTGGACCCAAGCCCTGAGGTACAGAAGAAGAA
GTATGCCTTCAAATGTCACAGACTAAAAGAAAATAATATTGAGCAGATTTACCCAGTCAATG

<210> SEQ ID NO 1176

<211> Length : 196

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1176
>H38804_PEA_1_node_24
ATATTTGGGATCCATTTAACAAAAAGCGACTGTGCCAATTCCATCGGTACCCCACGAGCATCGCATCACTTGCCTTC
AGTAATGATGGGACTACGCTTGCAATAGCGTCATCATATATGTATGAAATGGATGACACAGAACATCCTGAAGATGG
TATCTTCATTCGCCAAGTGACAGATGCAGAAACAAAACCCAA

<210> SEQ ID NO 1177
<211> Length : 353
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1177
>H38804_PEA_1_node_25
GTGAGTATGCTTCACCTGTATTTGAGCCTTTTCTTGCATTCAACCCAGGATTTATTAATTTTTCTAAATTCATGAAT
AGCATTGTTGATGCCTGCTCGATATTACAGCTGACTGTAGGGTTGGAGTTGATGTTATCATGTTCTCCCAAGCTTTC
AATATCCGTAGGTTGATAGACGTCTGATGGATAAAATTGTGCCTAGTTGTTTTGTAGAGAAGAATGTCAAACTCTTA
TTCTTCTTGAATAGGCTCTATTATTTGAATCTCTGGAGTTATTACCAGCTCATTGCTTCAAAATTAAGTTGAGGAAT
TCAAGAATAATTTATTTTAGTAAATTCTATTTAAGATGTTTAAGA

<210> SEQ ID NO 1178
<211> Length : 590
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1178
>H38804_PEA_1_node_28
TATTAACACAAAGTAAGTGACCTTCAGGTCTTATTGGAAACTCAGAGTAATATGGCCTTGCCTGGAATTGCAAATTT
CCTTAGTTTTGAAATTTTCATAGATGTCTTTGGTTCTTGGTTGTAACTGTTGACTGAGAAGAGCCATTTACATTTTT

TGATACCAACAGGGCAAAGCTTTTTACTTAATTACCTCTACCAGGCTTTAAGGGAAATCTGATACTTCAGCATGTGT
TAAACTATAAAATACCTACTCCAAGTATCTGCCCAGTTCCTTGTCCCCTCTCCCCAGGCCCTTAAAGGAAGTTCTCG
ATACATATTTGTAGAATAACTGAATGTTTTCAGGATTCCTGTACTTTGCTGAGTTAAAATGGATATGGTACCCTTGC
TGATTGGTTGAGCCCCTAAGAGGGGGCAGAATATTAAATATTCCATATCAGATATGCTTTTACAGGTTTGACTTTAG
AAAAGTCTTAGCATGTGAAGCCTGTTGGATAAAGGGCTGTGTTTGCATTTAATCTGTCACTTTTGTATCTCCTGTCC
TGGCTGGCCATTTTGATCTCATGCTGTTCTTTTTTTCTTTTGAACTTGTAG

<210> SEQ ID NO 1179

<211> Length : 1,228

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1179

>H38804_PEA_1_node_29

GTCACCATGTACTTGACAAGATTTCATTTACTTAAGTGCCATGTTGATGATAATAAAACAATTCGTACTCCCCAATG
GTGGATTTATTACTATTAAAGAAACCAGGGAAAATATTAATTTTAATATTATAACAACCTGAAAATAATGGAAAAGA
GGTTTTTGAATTTTTTTTTTTAAATAAACACCTTCTTAAGTGCATGAGATGGTTTGATGGTTTGCTGCATTAAAGGT
ATTTGGGCAAACAAAATTGGAGGGCAAGTGACTGCAGTTTTGAGAATCAGTTTTGACCTTGATGATTTTTTGTTTCC
ACTGTGGAAATAAATGTTTGTAAATAAGTGTAATAAAAATCCCTTTGCATTCTTTCTGGACCTTAAATGGTAGAGGA
AAAGGCTCGTGAGCCATTTGTTTCTTTTGCTGGTTATAGTTGCTAATTCTAAAGCTGCTTCAGACTGCTTCATGAGG
AGGTTAATCTACAATTAAACAATATTTCCTCTTGGCCGTCCATTATTTTCTGAAGCAGATGGTTCATCATTTCCTGG
GCTGTTAAACAAAGCGAGGTTAAGGTTAGACTCTTGGGAATCAGCTAGTTTTCAATCTTATTAGGGTGCAGAAGGAA
AACTAATAAGAAAACCTCCTAATATCATTTTGTGACTGTAAACAATTATTTATTAGCAAACAATTGATCCCAGAAGG
GCAAATTGTTTGAGTCAGTAATGAGCTGAGAAAAGACAGAGCATATCTGTGTATTTGGAAAAATAATTGTAACGTAA
TTGCAGTGCATTTAGACAGGCATCTATTTGGACCTGTTTCTATCTCTAAATGAATTTTTGGAAACATTAATGAGGTT
TACATATTTCTCTGACATTTATATAGTTCTTATGTCCATTTCAGTTGACCAGCCGCTGGTGATTAAAGTTAAAAAGA
AAAAAATTATAGTGAGAATGAGATTCATTTCAATGTAATGCACTAAAGCAGAACACGAACTTAGCTTGGCCTATTCT
AGGTAGTTCCAAATAGTATTTTTGTTGTCAAACTTTAAAATTTATATTAATTTGCAAATGTATGTCTCTGAGTAGGA
CTTGGACCTTTCCTGAGATTTATTTTATCCGTGATGTATTTTTTTTAATTCTTTTGATACAGAGAAGGGTCTTTTTT
TTTTTTAAGTATTTCAGTGAAAACTTGGTGTAAGTCTGAACCCATCTTTTGAAATGTATTTTCTTCATTGCAG

<210> SEQ ID NO 1180

<211> Length : 326

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1180

>H38804_PEA_1_node_30

GTCCACCTAATCATCCTGTGAAAGTGGTTTCTCTATGGAAAGCTTTGTTTGCTTCCTACAAATACATGCTTATTCCT
TAAGGGATGTGTTAGAGTTACTGTGGATTTCTCTGTTTTCTGTCTTACAAGAAACTTGTCTATGTACCTTAATACTT
TGTTTAGGATGAGGAGTCTTTGTGTCCCTGTACAGTAGTCTGACGTATTTCCCCTTCTGTCCCCTAGTAAGCCCAGT
TGCTGTATCTGAACAGTTTGAGCTCTTTTTGTAATATACTCTAAACCTGTTATTTCTGTGCTAATAAACGAGATGCA
GAACCCTTGAAAAATGGA

<210> SEQ ID NO 1181

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1181

>H38804_PEA_1_node_10

GATCCAACGCATGCCTGGAGTGGAGGACTAGATCATCAATTGAAAATGCATGATTTGAACACTGATCAAG

<210> SEQ ID NO 1182

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1182

>H38804_PEA_1_node_12

AAAATCTTGTTGGGACCCATGATGCCCCTATCAGATGTG

<210> SEQ ID NO 1183

<211> Length : 79

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1183
>H38804_PEA_1_node_13
TTGAATACTGTCCAGAAGTGAATGTGATGGTCACTGGAAGTTGGGATCAGACAGTTAAACTGTGGGATCCCAGAACT
CC

<210> SEQ ID NO 1184

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1184
>H38804_PEA_1_node_14
TTGTAATGCTGGGACCTTCTCTCAGCCTGAAAAG

<210> SEQ ID NO 1185

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1185
>H38804_PEA_1_node_2
ATGACCGGTTCTAACGAGTTCAAGCTGAACCAG

<210> SEQ ID NO 1186

<211> Length : 39

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1186
>H38804_PEA_1_node_20
CCATTTCTTTTCACAATATCCACAATACATTTGCCACAG

<210> SEQ ID NO 1187

<211> Length : 21

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1187
>H38804_PEA_1_node_23
GTGGTTCTGATGGCTTTGTAA

<210> SEQ ID NO 1188

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1188
>H38804_PEA_1_node_26
ATTTGAACTGCCAAAAATCTTTCCTCTCCACAGAGGTTGTTTCTTTAA

<210> SEQ ID NO 1189

<211> Length : 38

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1189
>H38804_PEA_1_node_3
CCACCCGAGGATGGCATCTCCTCCGTGAAGTTCAGCCC

<210> SEQ ID NO 1190

<211> Length : 111

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1190
>H38804_PEA_1_node_4
CAACACCTCCCAGTTCCTGCTTGTCTCCTCCTGGGACACGTCCGTGCGTCTCTACGATGTGCCGGCCAACTCCATGC
GGCTCAAGTACCAGCACACCGGCGCCGTCCTGGA

<210> SEQ ID NO 1191

<211> Length : 13

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1191
>H38804_PEA_1_node_5
CTGCGCCTTCTAC

Segment nucleic acid sequences:

<210> SEQ ID NO 1192

<211> Length : 257

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1192

>HSENA78_node_0

AGTGGGGAGAGATGAGTGTAGATAAAAGGAGTGCAGAAGGCACGAGGAAGCCACAGTGCTCCGGATCCTCCAATCTT

CGCTCCTCCAATCTCCGCTCCTCCACCCAGTTCAGGAACCCGCGACCGCTCGCAGCGCTCTCTTGACCACTATGAGC

CTCCTGTCCAGCCGCGCGGCCCGTGTCCCCGGTCCTTCGAGCTCCTTGTGCGCGCTGTTGGTGCTGCTGCTGCTGCT

GACGCAGCCAGGGCCCATCGCCAGCG

<210> SEQ ID NO 1193

<211> Length : 133

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1193

>HSENA78_node_2

CTGGTCCTGCCGCTGCTGTGTTGAGAGAGCTGCGTTGCGTTTGTTTACAGACCACGCAGGGAGTTCATCCCAAAATG

ATCAGTAATCTGCAAGTGTTCGCCATAGGCCCACAGTGCTCCAAGGTGGAAGTGGT

<210> SEQ ID NO 1194

<211> Length : 1,786

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1194

>HSENA78_node_6

TGGAAACAAGGAAAACTGATTAAGAGAAATGAGCACGCATGGAAAAGTTTCCCAGTCTTCAGCAGAGAAGTTTTCTG
GAGGTCTCTGAACCCAGGGAAGACAAGAAGGAAAGATTTTGTTGTTGTTTGTTTATTTGTTTTTCCAGTAGTTAGCT
TTCTTCCTGGATTCCTCACTTTGAAGAGTGTGAGGAAAACCTATGTTTGCCGCTTAAGCTTTCAGCTCAGCTAATGA
AGTGTTTAGCATAGTACCTCTGCTATTTGCTGTTATTTTATCTGCTATGCTATTGAAGTTTTGGCAATTGACTATAG
TGTGAGCCAGGAATCACTGGCTGTTAATCTTTCAAAGTGTCTTGAATTGTAGGTGACTATTATATTTCCAAGAAATA
TTCCTTAAGATATTAACTGAGAAGGCTGTGGATTTAATGTGGAAATGATGTTTCATAAGAATTCTGTTGATGGAAAT
ACACTGTTATCTTCACTTTTATAAGAAATAGGAAATATTTTAATGTTTCTTGGGGAATATGTTAGAGAATTTCCTTA
CTCTTGATTGTGGGATACTATTTAATTATTTCACTTTAGAAAGCTGAGTGTTTCACACCTTATCTATGTAGAATATA
TTTCCTTATTCAGAATTTCTAAAAGTTTAAGTTCTATGAGGGCTAATATCTTATCTTCCTATAATTTTAGACATTCT
TTATCTTTTTAGTATGGCAAACTGCCATCATTTACTTTTAAACTTTGATTTTATATGCTATTTATTAAGTATTTTAT
TAGGAGTACCATAATTCTGGTAGCTAAATATATATTTTAGATAGATGAAGAAGCTAGAAAACAGGCAAATTCCTGAC
TGCTAGTTTATATAGAAATGTATTCTTTTAGTTTTTAAAGTAAAGGCAAACTTAACAATGACTTGTACTCTGAAAGT
TTTGGAAACGTATTCAAACAATTTGAATATAAATTTATCATTTAGTTATAAAAATATATAGCGACATCCTCGAGGCC
CTAGCATTTCTCCTTGGATAGGGGACCAGAGAGAGCTTGGAATGTTAAAAACAAAACAAAACAAAAAAAAACAAGGA
GAAGTTGTCCAAGGGATGTCAATTTTTTATCCCTCTGTATGGGTTAGATTTTCCAAAATCATAATTTGAAGAAGGCC
AGCATTTATGGTAGAATATATAATTATATATAAGGTGGCCACGCTGGGGCAAGTTCCCTCCCCACTCACAGCTTTGG
CCCCTTTCACAGAGTAGAACCTGGGTTAGAGGATTGCAGAAGACGAGCGGCAGCGGGGAGGGCAGGGAAGATGCCTG
TCGGGTTTTTAGCACAGTTCATTTCACTGGGATTTTGAAGCATTTCTGTCTGAATGTAAAGCCTGTTCTAGTCCTGG
TGGGACACACTGGGGTTGGGGGTGGGGGAAGATGCGGTAATGAAACCGGTTAGTCAGTGTTGTCTTAATATCCTTGA
TAATGCTGTAAAGTTTATTTTTACAAATATTTCTGTTTAAGCTATTTCACCTTTGTTTGGAAATCCTTCCCTTTTAA
AGAGAAAATGTGACACTTGTGAAAAGGCTTGTAGGAAAGCTCCTCCCTTTTTTTCTTTAAACCTTTAAATGACAAAC
CTAGGTAATTAATGGTTGTGAATTTCTATTTTTGCTTTGTTTTTAATGAACATTTGTCTTTCAGAATAGGATTCTGT
GATAATATTTAAATGGCAAAAACAAAACATAATTTTGTGCAATTAACAAAGCTACTGCAAGAAAAATAAAACATTTC
TTGGTAAAAACGTAT

<210> SEQ ID NO 1195

<211> Length : 153

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1195
>HSENA78_node_9
ATATAATATATATTATATATTTAGCATTGCTGAGCTTTTTAGATGCCTATTGTGTATCTTTTAAAGGTTTTGACCAT
TTTGTTATGAGTAATTACATATATATTACATTCACTATATTAAAATTGTACTTTTTTACTATGTGTCTCATTGGTT

<210> SEQ ID NO 1196

<211> Length : 110

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1196
>HSENA78_node_3
GTAAGTTCTGTGCTGCTGTGTCCGCTGTGACCTTGGCAAGAGAGAAATCCCGCAGCCTGGGTCTTCAACCTTGGTAT
CTCATGAGTGTATCTTCTTTTTCTTTCCTTCAG

<210> SEQ ID NO 1197

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1197
>HSENA78_node_4
AGCCTCCCTGAAGAACGGGAAGGAAATTTGTCTTGATCCAGAAGCCCCTTTTCTAAAGAAAGTCATCCAGAAAATTT
TGGACGG

<210> SEQ ID NO 1198

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1198
>HSENA78_node_8
GTATTTATATATTATATATTTAT

Segment nucleic acid sequences:

<210> SEQ ID NO 1199

<211> Length : 139

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1199
>HUMODCA_node_1
GTGCGTCTCCATGGCGACCCGCCGGTGCTATAAGTAGGGAGCGGCGTGCCGTGGGGCTTTGTCAGTCCCTCCTGTAG
CCGCCGCCGCCGCCGCCCGCCGCCCCTCTGCCAGCAGCTCCGGCGCCACCTCGGGCCGGCGT

<210> SEQ ID NO 1200

<211> Length : 135

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1200
>HUMODCA_node_25
GTTGGTTTTGCGGATTGCCACTGATGATTCCAAAGCAGTCTGTCGTCTCAGTGTGAAATTCGGTGCCACGCTCAGAA
CCAGCAGGCTCCTTTTGGAACGGGCGAAAGAGCTAAATATCGATGTTGTTGGTGTCAG

<210> SEQ ID NO 1201

<211> Length : 163

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1201
>HUMODCA_node_32
ATCACCGGCGTAATCAACCCAGCGTTGGACAAATACTTTCCGTCAGACTCTGGAGTGAGAATCATAGCTGAGCCCGG
CAGATACTATGTTGCATCAGCTTTCACGCTTGCAGTTAATATCATTGCCAAGAAAATTGTATTAAAGGAACAGACGG
GCTCTGATG

<210> SEQ ID NO 1202

<211> Length : 215

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1202
>HUMODCA_node_36
AGACCTAAACCAGATGAGAAGTATTATTCATCCAGCATATGGGGACCAACATGTGATGGCCTCGATCGGATTGTTGA
GCGCTGTGACCTGCCTGAAATGCATGTGGGTGATTGGATGCTCTTTGAAAACATGGGCGCTTACACTGTTGCTGCTG
CCTCTACGTTCAATGGCTTCCAGAGGCCGACGATCTACTATGTGATGTCAGGGCCTGCGTG

<210> SEQ ID NO 1203

<211> Length : 173

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1203
>HUMODCA_node_39

GATGCCAGCACCCTGCCTGTGTCTTGTGCCTGGGAGAGTGGGATGAAACGCCACAGAGCAGCCTGTGCTTCGGCTAG
TATTAATGTGTAGATAGCACTCTGGTAGCTGTTAACTGCAAGTTTAGCTTGAATTAAGGGATTTGGGGGGACCATGT
AACTTAATTACTGCTAGTT

<210> SEQ ID NO 1204

<211> Length : 1,096

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1204

>HUMODCA_node_41

TTGAATATTTGTTTTATATGGATTTTTATTCACTCTTCAGACACGCTACTCAAGAGTGCCCCTCAGCTGCTGAACAA
GCATTTGTAGCTTGTACAATGGCAGAATGGGCCAAAAGCTTAGTGTTGTGACCTGTTTTTAAAATAAAGTATCTTGA
AATAATTAGGCATTGGGACGTTTTTATGGTGTGTTCATTCCAGACAGTTCACGAATCCCGTATAGCTCGCTCTGATT
CTCAGAGAACAATGAGTGGGTCCACCCACACACAGGTAGGAGGACAGGTGAGACGGAAGCCCCATCCTCCCATGTGG
ACGGTGCACATCTGCTCAGCCCACCCCACATGTCCAGAGTTGGCTGCAAACTCCTTGTCCAGAGCCTCTGGTGGTGG
GACCTACTTAAGTCTGACGGACCTGTCCTGTCCAGGCCAGTGCCCAGGGAAGGTGTGGGAGGCCCTTTGAGCCTGGC
CTGCAGAGACCATCCGTGTCCCCTCCCACCTTCATGCCTGTGAGAAGTTAGGAATGTATACGGTACCACATTTGGCA
GTCAGCTTATTTTAATAAATTCAGCAACAGCAAGTCCCTACCATGTTGTGTATCTTCACCATCTTGTCTGACCATGA
CCACTGGCCTTGTGTGTTCTTTTACTCAACGTGTACCCCCGCTCTCCCCCAAAGTGTGGCAGGCTCTCATGCTCCTT
AACCCCCATTGTGGCAATGTCTTACGGGTAACGCTGGAGCTGCAGGAGGAGGGAAGGACACGTCAGAGCCACCAGGC
AGTGGGAGCATCTTGGAGTCCCCACCAGCCTCATGAGGGGGACAGGAAGAGAGCAAATGTGTAGGGAGGAAGGCTGT
GGCTCCTCCCGGGGTGGGAAGGTCAAGCCGATGCTGTCACCCATTTACCAAAGCTGAAGAGAGTGACTTCCTTTCTC
AAAAGCATCACCTTCCCCTGAACCCTGAGTCCAGAGAAGCCAGGAGCCCTCATGTGGCTGCCGAGTTAGCTCAGGGC
TTGGCTTATCACCAACTCTGGTCTCCCTGGGCCAGGGTTGCCAAAACATGAAAGATTTTTTCAGGAGCCAGAGGTTG
GTTCTGACTGGAGGGGGA

<210> SEQ ID NO 1205

<211> Length : 117

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1205
>HUMODCA_node_0
GACGTCGGCCCGCCGGCGCCCCACCAGCTCCGCGCGGGCCCGGGTTGGCCACCGCCGGGCCCCCGCCCCTCCCCCGG
CGGTGTCCCGGCCGGAACCGATCGTGGCTGGTTTGAGCTG

<210> SEQ ID NO 1206
<211> Length : 110
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1206
>HUMODCA_node_10
ATTGTCACTGCTGTTCCAAGGGCACACGCAGAGGGATTTGGAATTCCTGGAGAGTTGCCTTTGTGAGAAGCTGGAAA
TATTTCTTTCAATTCCATCTCTTAGTTTTCCAT

<210> SEQ ID NO 1207
<211> Length : 92
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1207
>HUMODCA_node_12
AGGAACATCAAGAAATCATGAACAACTTTGGTAATGAAGAGTTTGACTGCCACTTCCTCGATGAAGGTTTTACTGCC
AAGGACATTCTGGAC

<210> SEQ ID NO 1208
<211> Length : 27
<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1208
>HUMODCA_node_13
CAGAAAATTAATGAAGTTTCTTCTTCT

<210> SEQ ID NO 1209

<211> Length : 72

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1209
>HUMODCA_node_2
CTCCGGCGGGCGGGAGCCAGGCGCTGACGGGCGCGGCGGGGGCGGCCGAGCGCTCCTGCGGCTGCGACTCAG

<210> SEQ ID NO 1210

<211> Length : 82

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1210
>HUMODCA_node_27
CTTCCATGTAGGAAGCGGCTGTACCGATCCTGAGACCTTCGTGCAGGCAATCTCTGATGCCCGCTGTGTTTTTGACA
TGGGG

<210> SEQ ID NO 1211

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1211
>HUMODCA_node_3
GCTCCGGCGTCTGCGCTTCCCCATGGGGCTGGCCTGCGGCGCCTGGGCGCTCTGAG

<210> SEQ ID NO 1212

<211> Length : 84

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1212
>HUMODCA_node_30
GCTGAGGTTGGTTTCAGCATGTATCTGCTTGATATTGGCGGTGGCTTTCCTGGATCTGAGGATGTGAAACTTAAATT
TGAAGAG

<210> SEQ ID NO 1213

<211> Length : 113

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1213
>HUMODCA_node_34
ACGAAGATGAGTCGAGTGAGCAGACCTTTATGTATTATGTGAATGATGGCGTCTATGGATCATTTAATTGCATACTC
TATGACCACGCACATGTAAAGCCCCTTCTGCAAAAG

<210> SEQ ID NO 1214

<211> Length : 55

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1214
>HUMODCA_node_38
GCAACTCATGCAGCAATTCCAGAACCCCGACTTCCCACCCGAAGTAGAGGAACAG

<210> SEQ ID NO 1215

<211> Length : 94

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1215
>HUMODCA_node_40
TTGAAATGTCTTTGTAAGAGTAGGGTCGCCATGATGCAGCCATATGGAAGACTAGGATATGGGTCACACTTATCTGT
GTTCCTATGGAAACTAT

Segment nucleic acid sequences:

<210> SEQ ID NO 1216

<211> Length : 271

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1216
>R00299_node_2
GCGGCCGCAGAGCACTTTGCCCGGAGCCCAGCGTCCTCCCCTGAGTTCGCTGAGTCTCCCGGGACCAGCAAAGGCTG
CGCGCCCCGCATCGGCCCGGAGGCGGGGAGCCCTGGGAGGCCTGGCCGAGCTGCCCGCAGGGAAATGGCGGAGAAAG
CGCTTCTCTGCCCGAGTTCAGCCGGGCTGGGGACTTGGCCCTGGGTCCTGAACTCGGCATGGCCAGTTCTGCCTCTG
GCTGTGGACCAGGGTGTGGACTGGAGACCGCGGGGGCCAG

<210> SEQ ID NO 1217

<211> Length : 172

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1217
>R00299_node_30
GGGATCGACATTGAGACCAAGATGCACGTCCGCTTCCTTAACATGGAAACCATGGCCCTCTGCCACTGACCCACCGC
CACCTCCGCGGAGAAACTGCACTTTGCAATGGGGCCGCCTCCCCGCGTAGCTGGAGCAGCCCAGGCCCGGCGGACAG
CCTCTTCCTGCAGCGCCG

<210> SEQ ID NO 1218

<211> Length : 77

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1218
>R00299_node_10
GAGAACTTCAACAATGTCCCGGACCTGGAGCTCAACCCCATCCGATCCAAAATTGTTCGTGCCTTCTTCGACAACAG

<210> SEQ ID NO 1219

<211> Length : 115

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1219
>R00299_node_14

GAACCTGCGCAAGGGACCCAGTGGCCTGGCTGATGAGATCAATTTCGAGGACTTCCTGACCATCATGTCCTACTTCC
GGCCCATCGACACCACCATGGACGAGGAACAGGTGGAG

<210> SEQ ID NO 1220

<211> Length : 25

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1220
>R00299_node_15
CTGTCCCGGAAGGAGAAGCTGAGAT

<210> SEQ ID NO 1221

<211> Length : 62

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1221
>R00299_node_20
TTCTGTTCCACATGTACGACTCGGACAGCGACGGCCGCATCACTCTGGAAGAATATCGAAAT

<210> SEQ ID NO 1222

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1222
>R00299_node_23

GTGGTCGAGGAGCTGCTGTCGGGAAACCCTCACATCGAGAAGGAGTCCGCTCGCTCCATCGCCGACGGGGCCATGAT
GGAGGCGGCCAGCGTGTGCATGGGGCAGATG

<210> SEQ ID NO 1223

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1223
>R00299_node_25
GAGCCTGATCAGGTGTACGAGGGGATCACCTTCGAGGACTTCCTGAAG

<210> SEQ ID NO 1224

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1224
>R00299_node_28
ATCTGGCAG

<210> SEQ ID NO 1225

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1225
>R00299_node_31

GTACATAGCCAAGGCTCGTCTGCGCACCTTGTGTCTTGTAGGGTATGGTATGTGGGACTTCGCTGTTTTTATCTCCA
ATAAAAAAAAAAAAAAGGTTTGTTAATTAAT

<210> SEQ ID NO 1226

<211> Length : 70

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1226
>R00299_node_5
TCTCATCGGATCAGATCGAGCAGCTCCATCGGAGATTTAAGCAGCTGAGTGGAGATCAGCCTACCATTCG

<210> SEQ ID NO 1227

<211> Length : 4

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1227
>R00299_node_9
CAAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1228

<211> Length : 157

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1228
>W60282_PEA_1_node_10
GGCTTGTAGGGGGAGAGACCAGGATCATCAAGGGGTTCGAGTGCAAGCCTCACTCCCAGCCCTGGCAGGCAGCCCTG
TTCGAGAAGACGCGGCTACTCTGTGGGGCGACGCTCATCGCCCCCAGATGGCTCCTGACAGCAGCCCACTGCCTCAA
GCC

<210> SEQ ID NO 1229

<211> Length : 137

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1229
>W60282_PEA_1_node_18
TACGCCTGCCTCACACCTTGCGATGCGCCAACATCACCATCATTGAGCACCAGAAGTGTGAGAACGCCTACCCCGGC
AACATCACAGACACCATGGTGTGTGCCAGCGTGCAGGAAGGGGGCAAGGACTCCTGCCAG

<210> SEQ ID NO 1230

<211> Length : 436

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1230
>W60282_PEA_1_node_22
TCTCTTCAAGGCATTATCTCCTGGGGCCAGGATCCGTGTGCGATCACCCGAAAGCCTGGTGTCTACACGAAAGTCTG
CAAATATGTGGACTGGATCCAGGAGACGATGAAGAACAATTAGACTGGACCCACCCACCACAGCCCATCACCCTCCA
TTTCCACTTGGTGTTTGGTTCCTGTTCACTCTGTTAATAAGAAACCCTAAGCCAAGACCCTCTACGAACATTCTTTG
GGCCTCCTGGACTACAGGAGATGCTGTCACTTAATAATCAACCTGGGGTTCGAAATCAGTGAGACCTGGATTCAAAT
TCTGCCTTGAAATATTGTGACTCTGGGAATGACAACACCTGGTTTGTTCTCTGTTGTATCCCCAGCCCCAAAGACAG
CTCCTGGCCATATATCAAGGTTTCAATAAATATTTGCTAAATGAGTGAATC

<210> SEQ ID NO 1231

<211> Length : 669

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1231

>W60282_PEA_1_node_5

GGAGCGGCCTAGGGGAGGCCAGGGGCCCACCTGGGCTGGGGCTGTGGAGAGGGAGTGGCTGGGACGGGAGGAAAAAG
AGAGACGGAGATTAGATGGAAGAAGAGGGATTTCAAGACAAATTGCCAGAGATGCAGTCAGAGAGACTGACTGAGAG
ACACAAAGATAGAAGGAATTAGAGAAAGGGCCACACAGAGCCAGACAGAGAGAGAAGAGTGGAGATGGAGACAGGGA
CGAGGACAGAGAAAGGCAGACAGACACATAGGGACAGAAAGAGAAAAATCACACAAAGTCAGAATTACTGAATGACA
GGGAATGACACATAGAACGAGACACAGATTCAGAGACTCAGGGCAGGGAAAGGAAGGCTGCAGACAGACAGACAGAC
AGAGGGAGGCTGAGACACAGGGAGAAGAGGGGCTTGGAGAGGTGGCACAGGCAGGCAGCCAGTGCCTCAGAGGCCTC
CGGGGAGGGCCCTCACACACACCCCGCCCCGGGGCATTAAGGCAGGGCTTGGAGGCCAGTCATCCTGGGCCCGCCCA
GGGCCGCCCCCCTGCCAGCCCGCCTGCCTGGTGCCTGGCACCTGGCGCTCCAACCCAGCCTACCTGCTGTAGCTGCC
GCCACTGCCGTCTCCGCCGCCACTGGGCCCCCAGAGCCCCAGCCCCAGAGCCT

<210> SEQ ID NO 1232

<211> Length : 33

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1232

>W60282_PEA_1_node_21

GGTGACTCCGGGGGCCCTCTGGTCTGTAACCAG

<210> SEQ ID NO 1233

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1233
>W60282_PEA_1_node_8
AGGAACCTGGGGCCCGCTCCTCCCCCCTCCAGGCCATGAGGATTCTGCAGTTAATCCTGCTTGCTCTGGCAACAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1234

<211> Length : 616

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1234
>Z41644_PEA_1_node_0
CCTTGCTGTTCATGGCCCAGCAGGGGCCCTATGGGGGTCAGGCCTGCAGGCACTCACACTTGGCACCTGCTCCAAAA
CCCTTTCAGGTCTTTGAGGATCTGAGCCCTGGGCCTGGGTCTCCCGCCGGCTCGGAAAAGCTGGCCTGCCGGGCCAG
ACGAGAGAACCACACGATTCAGAAAAGCAGTGCCCTTCAGCAGCCTCTCCACCGTCTGGGCTCCCCAAAGGCAGAGC
GGGACGCTGGAAATGTGTGCGCGCTGTGGTATGGGTGTGCAAGTGTGCGAAGGCGGCGTGTTGTGTGAGCGAGAGGG
TAGCGGATGTGTGTGTGCGTGTGCGCGCGTGGCTCCGGGTGTGCGCCGCTGCGATAGCGGGTCCTTTCCCGGGGCGG
GCGACGGGCGGGCTGGGAAGGTCTCCTCCCCTCACCACATTGAGAAATCTCAGTGAGTCACCGAGTGGTTCTGCATA
TTAATGAGCTCGCTCGCTGCGAGGGCAGGAGCGGATTTAAAAGAGGCCAGGGCGGGCGGAGGGAGGCTGTGGAGAGA
GCGCGGAGACAAGCGCAGAGCGCAGCGCACGGCCACAGACAGCCCTGGGCATCCACCGACGGCGCAGCCGGAGCCAG

<210> SEQ ID NO 1235

<211> Length : 1,062

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1235
>Z41644_PEA_1_node_11

GTACGCCCCGCCTCTACTCACCTTCCTTCCCACCAGACCCAGCTGTGGCTCTCAGGATGGGAAGGGACCCCCCCACC
AGGTCATCTAGCCCCATCTAATATGTGAACACCCACCACAACATCCACAGCAAACAGATACTCAGACAATGCTTACA
TACCCCCAGGGACAAGGAACTCACCACTTACCAAAGCTAGCTATCCATCTCTTGTCCATTTGCAAGCATGGCAGGTT
TGTCATTTTGTAAACTAAAGTCTGTCTCACTCTAATATTTGCATTATAATCTTAATTCCTCTTTTTATTTCAGTTAC
GTAAGTTGTTAAATGGCAGAGTGAGCACTGGCATGGCTGCCAGGGGAGCTCTGAGGACTTCAGTGGGGTGAAATGTG
ACCACTTAGGTGACTGTGTATGTTGGCTATAAAACTGCGCTATAAAACCATGAGGTGCTGAGGATGATCCTTGCCAG
AAACATGTTTTCTTCTCCAAGGTGCCCCACTCCCTCTGCTGCCCAGAAACCTGATAAACTCCTTCCTTCGCAGGTGC
TGGAAGGCACCACAGGTTTGGCTCTTTAAAATCAGAGCCACTGTTAACCAAGGCGGGCAGCAGTGTTAAGACCACCA
GCACCCTGAACCAGCCCTGTACTTACTGGGCACTGTTTCCTTAAAATCAGAAGGTGGCTTCCCATCTCTGGTTTCCT
GGGGTCTTATGTCTGTCCTCGGAGGGAGAATCCAGTTCCTAGCTCCCCTGTACCATGCGAAGGTAGCCTGTCCTGTC
TCACTCCTCAGATACGCAGAGTCTGTTTACACATTTGCCTGCATAGCATGATCAGGAAGCACACACACACACACACA
CACACACACACACGCATGCATGCACACACCATGCAGGTGACTTCCCCAGGAACTAGTGCCAGCACCCCTGCTGCAGA
GGGGGATATCAAGGCTAAATGGAAGAGAGGGGTGACTTGCCTGGGAGCACAGGGCAAAGCCAGGACAGCAAACCAGG
CCTCCTGGTGCTACCCCACCAGCTGCCCTCACAGGGTGGAAGGTACAGCCATAGTGGGTGC

<210> SEQ ID NO 1236

<211> Length : 261

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1236

>Z41644_PEA_1_node_12

CTGCATTGCCCTCCCCTCACCTGGCCCAGCCATGCTACCCCAAGCTCAGCCCTGTGACCAGCTCTCCCAGAGCTGAC
ACTCGGGCTCAACCCCTATACCTGAGCCTTTTTTGCTGCCTCCAAAACAGCCTCATCTGCAGTTGCTTGAAATAGAA
AGTGATGAGAGCAATAAATTATTTTCTATAAATCTGCTGGGAATGAAGCCCTCTTTCTGGTCAAGCCAGGCAGCTCA
TGTGGCAAAGGCCAGAACTGCGCAGTCCAC

<210> SEQ ID NO 1237

<211> Length : 1,361

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1237

>Z41644_PEA_1_node_15

GCCCTGTGTCATCAAAACTGGCAGAAGGCTAATCCCATGGGCAGGTTATGGAAGAGGCTGAGGGCATCTTGATCTGA
TTGCTGGGGGATACTCAAACCTTTAGCTCACCTTGCTTCTCCCCTCCACCTGAGCTGCAGCCTGGAAAAGGAGGCAC
CCACAGGTCTAAACATGGCCCTGCTTTTTTTTTTCCTGAAAATTCCAATAACAAAAGCAACGAGAGCCTCTCACTAC
CAGGCCTTCTCTCACTTTGCTATAAAATTAGTTCACCCCTCTTTCTTAGAGTGTTGAGGTCCCTGCCCTCCCCACCT
CCCTCCCCTGAAACAAGTTGAAAATATCTTAATGAACATAGAACAGTGATAAAGGAAGTGTTTGAAGTCCTCTTTGT
ACAGAGAGAGAGAGAAAGAGAATGCCAAAGCTAGGTTGGAGGAAGTAGAAGGGTATACGGTGGGCTCAGGCCCATGG
GGGCCACACAGAGGAGCTCTGTGCACTTCAGAGACCAGAGCTTCCAGGGAGCTTCTGGCCACCACAGGAAGCAGCCT
AGTCAGGCATTTTATTTCAATGGATAATTCAGTGGTCTTACTCAGAAATCAAGAACGAGACAGAAAAGTGATAGGCT
AAGTGTAACGTATGGCCCCAGGGCAGCCATGGGGCAGAACTAGAAGAAAGCAAAATATCTAACTGGGCACAGCTTGA
GAGGTGAGGGGAAGGTGGGGCTGGGAACGAGTAGAGATGAGGCAATGCAGCCAGGAGCAGGGACTGAGGGGCACAGG
CCTCCTGCACCACTGCCCCACCCCACCAACCACCTCTTCTGTCTCCAGGAAGCAGCTTCTAGAGCTAGCATTCTTCT
GGAGGACATGCATTATTTGGGCAAAATACAAAGAAATATACAAGCCTAAGTCAAGTAAGGGAATGCCTCCCACCCTT
GCTATTTTCTCTAAATAGAGAGGCTGAGTACAGACGCGGAAAGAAACAAGGAGGTGTGGGAGCAGCCCGCCATGCTA
GAGAAAGACTACATTCCTGCCACTAACAGTCGGTGGCCACTGGGCAAATCTTAAGTCTGTGGTGCCTCAGTTTCCTC
ATATGCAAAGCGGGTTTGTTCCATAGGCCTCTGAGGACAAAATGAGATTGCAGAAGTGAGATTGCAGATGGTTAGAA
AAGACAAAGCCACACTGGTGTGAGTTTTCATGGTCCCCGGGACCACATCCTCAGAAGGATCCCTCCCACTTCTCCTG
GGGGTTCCTGCAGTTCTGGGACAGGGGCATTCCCTGCAGACCAGACGTGAATGAAGCCGCTTAGCCAGCATCTTGTG
AACGGCCTGCCTCATGTCCTGAGCCACTTACACATGTGTTTTTTCTCCCCAG

<210> SEQ ID NO 1238

<211> Length : 569

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1238

>Z41644_PEA_1_node_20

ATGGGAGACCCATCTCTCTTGTGCTCCAGACTTCATCACAGGCTGCTTTTTATCAAAAAGGGGAAAACTCATGCCTT
TCCTTTTTAAAAAATGCTTTTTTGTATTTGTCCATACGTCACTATACATCTGAGCTTTATAAGCGCCCGGGAGGAAC
AATGAGCTTGGTGGACACATTTCATTGCAGTGTTGCTCCATTCCTAGCTTGGGAAGCTTCCGCTTAGAGGTCCTGGC

GCCTCGGCACAGCTGCCACGGGCTCTCCTGGGCTTATGGCCGGTCACAGCCTCAGTGTGACTCCACAGTGGCCCCTG
TAGCCGGGCAAGCAGGAGCAGGTCTCTCTGCATCTGTTCTCTGAGGAACTCAAGTTTGGTTGCCAGAAAAATGTGCT
TCATTCCCCCCTGGTTAATTTTTACACACCCTAGGAAACATTTCCAAGATCCTGTGATGGCGAGACAAATGATCCTT
AAAGAAGGTGTGGGGTCTTTCCCAACCTGAGGATTTCTGAAAGGTTCACAGGTTCAATATTTAATGCTTCAGAAGCA
TGTGAGGTTCCCAACACTGTCAGCAAAAAC

<210> SEQ ID NO 1239

<211> Length : 163

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1239
>Z41644_PEA_1_node_24
CAATATATTTGTGATTCCCCATGTAATTCTTCAATGTTAAACAGTGCAGTCCTCTTTCGAAAGCTAAGATGACCATG
CGCCCTTTCCTCTGTACATATACCCTTAAGAACGCCCCCTCCACACACTGCCCCCCAGTATATGCCGCATTGTACTG
CTGTGTTAT

<210> SEQ ID NO 1240

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1240
>Z41644_PEA_1_node_1
CAGAGCCGGAAGGCGCGCCCCGGGCAGAGAAAGCCGAGCAGAGCTGGGTGGCGTCTCCGGGCCGCCGCTCCGACGGG
CCAG

<210> SEQ ID NO 1241

<211> Length : 56

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1241

>Z41644_PEA_1_node_10

CTGCAGAGCACCAAGCGCTTCATCAAGTGGTACAACGCCTGGAACGAGAAGCGCAG

<210> SEQ ID NO 1242

<211> Length : 17

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1242

>Z41644_PEA_1_node_13

ACTCTGTCACCCTCCAG

<210> SEQ ID NO 1243

<211> Length : 81

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1243

>Z41644_PEA_1_node_16

GGTCTACGAAGAATAGGGTGAAAAACCTCAGAAGGGAAAACTCCAAACCAGTTGGGAGACTTGTGCAAAGGACTTTG

CAGA

<210> SEQ ID NO 1244

<211> Length : 20

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1244
>Z41644_PEA_1_node_17
TTAAAAAAAAAAAAAAAAAA

<210> SEQ ID NO 1245

<211> Length : 108

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1245
>Z41644_PEA_1_node_19
AAGCCTTTCTTTCTCACAGGCATAAGACACAAATTATATATTGTTATGAAGCACTTTTTACCAACGGTCAGTTTTTA
CATTTTATAGCTGCGTGCGAAAGGCTTCCAG

<210> SEQ ID NO 1246

<211> Length : 40

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1246
>Z41644_PEA_1_node_2
CGCCCTCCCCATGTCCCTGCTCCCACGCCGCGCCCCTCCG

<210> SEQ ID NO 1247

<211> Length : 23

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1247
>Z41644_PEA_1_node_21
CTTAGGAGAAAACTTAAAAATAT

<210> SEQ ID NO 1248

<211> Length : 53

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1248
>Z41644_PEA_1_node_22
ATGAATACATGCGCAATACACAGCTACAGACACACATTCTGTTGACAAGGGAA

<210> SEQ ID NO 1249

<211> Length : 54

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1249
>Z41644_PEA_1_node_23
AACCTTCAAAGCATGTTTCTTTCCCTCACCACAACAGAACATGCAGTACTAAAG

<210> SEQ ID NO 1250

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1250
>Z41644_PEA_1_node_25
ATGCTATGTACATGTCAGAAACCATTAGCATTGCATGCAGGTTTCATATTCTTTCTAAGATGGAAAGTAATAAAATA
TATTTGAAATGTACCAAAATTCTAGA

<210> SEQ ID NO 1251

<211> Length : 36

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1251
>Z41644_PEA_1_node_3
GTCAGCATGAGGCTCCTGGCGGCCGCGCTGCTCCTG

<210> SEQ ID NO 1252

<211> Length : 34

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1252
>Z41644_PEA_1_node_4

CTGCTGCTGGCGCTGTACACCGCGCGTGTGGACG

<210> SEQ ID NO 1253

<211> Length : 106

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1253
>Z41644_PEA_1_node_6
GGTCCAAATGCAAGTGCTCCCGGAAGGGACCCAAGATCCGCTACAGCGACGTGAAGAAGCTGGAAATGAAGCCAAAG
TACCCGCACTGCGAGGAGAAGATGGTTAT

<210> SEQ ID NO 1254

<211> Length : 58

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1254
>Z41644_PEA_1_node_9
CATCACCACCAAGAGCGTGTCCAGGTACCGAGGTCAGGAGCACTGCCTGCACCCCAAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1255

<211> Length : 669

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1255
>Z44808_PEA_1_node_0
CCTGGACCCTGGGGCGTGAGGAGGGCGCGGTGCGTCCCGTGGTTGTGCTTGGAAGCCCCCCGAGGGTGCGCGCGCGT
GGGTATGAGTGCGTGCGTGTGCCTGGGTGTGCGTGTGTGTAAGTGTGCACGTGTGTGTGTGAGAGTGCGCGCGGGGA
AGGAGGCACAGAGACAGCCCGGACAGGCCACTGCGCAGCCCTGGTGGCCCCCGCTCCACCTCTCGCTCCGCAGACCC
GCGCCAGGGAGGCCTCTGGGCCGCAGCGGGCACCGGAGCGGAGCGGGCGCGGCAGCGGGCGCTGGGAGGTGGGGCTG
GGGGAGGAGAGGGGGAGGGAGAGAGGCGGGCGGGAGGGGAGGATCCGGGAAGCTCCGGGGTATTTGACAGGAGCGAG
GGCGGACGCAAAGAACGCGGAGGACCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGG
GAGAGCATCGCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCA
GGGCGCAGGACGCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACCATGCTGCTCCCCCAGCTCTGCTGGCTGCCGC
TGCTCGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGTTCTCGGCGCTCACG

<210> SEQ ID NO 1256

<211> Length : 187

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1256
>Z44808_PEA_1_node_16
TGTCATCCTGTGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACAATGTGGTGATCCCT
GAGTGTGCGCACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTACTGCTGGTGCGTCCTGGTGGA
CACGGGGCGCCCCATTCCCGGCACATCCACAAG

<210> SEQ ID NO 1257

<211> Length : 172

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1257
>Z44808_PEA_1_node_2
TTTTTGAGAGTGGATCAAGATAAAGACAAGGATTGTAGCTTGGACTGTGCGGGTTCGCCCCAGAAACCTCTCTGCGC
ATCTGACGGAAGGACCTTCCTTTCCCGTTGTGAATTTCAACGTGCCAAGTGCAAAGATCCCCAGCTAGAGATTGCAT
ATCGAGGAAACTGCAAAG

<210> SEQ ID NO 1258
<211> Length : 275
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1258
>Z44808_PEA_1_node_24
GCTCTCAGAACCCGACCCCAGCCATACCCTAGAGGAGCGGGTGGTGCACTGGTACTTCAAACTACTGGATAAAAACT
CCAGTGGAGACATCGGCAAAAAGGAAATCAAACCCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGT
GTGAAGAAGTTTGTTGAATACTGTGACGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGG
CGTGGCGAAAGAGGACGGCAAAGCGGACACCAAGAAACGCCACA

<210> SEQ ID NO 1259
<211> Length : 1,685
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1259
>Z44808_PEA_1_node_32
GATGCAATGGTGGTGTCCTCCAGACCCAAAGCCACAACCCATCGCAAGTCAAGAACACTTTCCAGAAGATAAACATG
AGTGGGTTCATGTCTCTCTCCTTCAAAGCCAGGACAAAATCCCCACTTCTTTGCTGCCGCGAGTCAATTTGTGATTT
ATTTTGTCTGCACCTGTTTGATGCCAGGTCGACATTTCCTAAGGCAAGCCCCTGTATTTGTTGTGGATTTAAGTGGA
GGCGGCCAGCACACACCTTGGATGTAATTTAAAACCATTTCCTGAGGAAAGATGTGTGATATGCTTTCCTTTGTTTA
GCAAATGTTTATGGTTTTAACTTTAAATCTCACCGCAAATCACTTACACTTGAAAACAGGGCTGGTCTGAAAGTAAT

TACCCTCCCTGAGTGCCAAGACCTCCAGAAGTTGTTTTCATTCCCGAATGGCAATCACTGTACTCATGCGCTCCACG
CATCTTAAATAAACTCAGTTCAAAGCACATGCCTCCTGCTTCAGCTCTTTTTTCCAAAAAGAGAAACAGAAGCAGGT
TCCCCCTCCTTTTATAGTGCCTGCGTGGACACGCGGACCTCCATGCCTTTCATGCTGTGGCTATGTCAGCAAACTAC
GATATTGGGATGATCCTAACGGGCAAGCCAGCTGCGGCTCCTACCGGCCGTGGCCATTGAAGGCCACCATGTTGCTT
TGAAACATCTCAAAGAATAACATAGTGCCAGCCAGCAAGGGTTTCACCATATGCATGACCCAGACAGGAACTATCAA
AAGAAGGGATCACGGGAAGGTGCATGATGCTAATGTGGAATCCAGAGGAGCTCTTTCCTGATCTCTTCAGCTTCCGC
TGCCACTCCAGAATCATCAGAGCTGATATTAAATAAGTTAAAATGTTAGTCCACCGTCTCCTCCTGCAATCCTAACC
ATCTTTTGAGACTGTTAGAATACTTTGACGGGTTGTCTTTCTGTGCAACTAATTTAAACCTCAAGTTTAGTGTAGGA
GATGGGTTTGTCTTCTCACCTCTTCAGATCTTTATCAAGGGGGAATAAAAGCCAACCCAGAAACCTAAACTTTAAAA
TTTAATTATTTGAAATAATAAAACAGAAGAAGGGATCAACATTTGTCGGAATTGGCACTCTTGGAAAACTAAGTCTA
GGAGATCATATATTGCTTTTTTTTTTTCATTCTAAATTACTTTTAATTGAAAGTCAAGATGCTGAGTTACAGTTGTT
TATCATTATAATAAGCAAACTTTTTAAGTTGGATTTCTTCTTAAAGAGGTAAACTAGTGAACAAAAAAGATAAAAAG
GAAAATTAAGAATCAACTATGCCTTTATCAAATTTGAAGCATAAGTTATATTATTAAAATTATTTTTGTATAATCAA
GGTGATAAGACATTCTGGAAAACATTTAATGTATTTAGTACTTAGAATATTTACAGTGGATGTTACTTTTTTGAAAC
GATATATTTTTCCCAATTTTTCTATCATGTCAAGGAAGGAAACTGTTAAGAAGTTACCAGTGTCCAAAATGTCTTCA
TTGTTTCTTACTCATACTTACACCTCACATGACCTGCCCAGCCCTCTTTGGTTCAGTTCATTCCCAGAAGCCAAGCC
TTAGTCTTCACAGATGAGCGACACACACCTCTGAATATAATGTCTCTTTTTTGTTTTTTCCTTTTCAG

<210> SEQ ID NO 1260

<211> Length : 877

<212> Type : DNA

<213> Organism : Homo sapiens

.

<400> sequence : 1260

>Z44808_PEA_1_node_33

CCAAGGAAACAAGGATAAATGGCTCATACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCACCAAAGAGC
AATTAAGAAAACAAAAACAGAAACACATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGAG
CTATTTAAACAGACTGTTTTAATCTGTGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAATGTATGTGTC
CTCTTTTGACCTTGGAAATCTGTATGTGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCGCCTTCCATA
TGTTAACAGTAGAGCTCTATGCACTCCGGCTGCAATCGTATGGCTTTCTCTAACCCCTGCAGTCACTTCCAGATGCC
TGTGCTTACAGCATTGTGGAATCATGTTGGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCCGACCCTACA
GGCAGTTAACATGCATGGGCTGGTTTGTTTCTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTTCCAGAGATCTT
GCTCATACAATGAATCACGCAACCACTAAAGCTATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAATAATATTTT
CAAACTGTCGTTGGTGTGATACTTTGGCTCAAAGGATCTTTGCTTTTCCATTTTAAGCTTCTGTTTTGAGTTTTGCC

CTGGGGCTTGAATGAGTCCCAGAGAGTCGTTCGGATGGTGGGAGGCTGCCTAGGAGGCAGTAAATCCAGTCACAGTG
CCTGGGAGGGGCCCATCCTTCCAAAATGTAAATCCAGTCGCGGTGTGACCGAGCTGGCTAACAGGCTTGTCTGCCTG
GTTTTCCTCCTACACGTGGACATTATTCTC

<210> SEQ ID NO 1261

<211> Length : 252

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1261
>Z44808_PEA_1_node_36
GATACGTGGTGCCCCCGGGGCTGGTGTTGGCAGCCGGGGGGAGGTGCCTGAGGGTCCCCACGGTTCCTTTCTGCTTT
TCTGAATGCATCAAGGGTACGAGAACTTGCCAATGGGAAATTCATCCGAGTGGCACTGGCAGAGAAGGATAGGAGTG
GAATGCCCACACAGTGACCAACAGAACTGGTCTGCGTGCATAACCAGCTGCCACCCTCAGGCCTGGGCCCCAGAGCT
CAGGGCACCCAGTGTCTTAAG

<210> SEQ ID NO 1262

<211> Length : 349

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1262
>Z44808_PEA_1_node_37
GAACCATTTGGAGGACAGTCTGAGAGCAGGAACTTCAAGCTGTGATTCTATCTCGGCTCAGACTTTTGGTTGGAAAA
AGATCTTCATGGCCCCAAATCCCCTGAGACATGCCTTGTAGAATGATTTTGTGATGTTGTGATGCTTGTGGAGCATC
GCGTAAGGCTTCTTGCTTATTTAAACTGTGCAAGGTAAAAATCAAGCCTTTGGAGCCACAGAACCAGCTCAAGTACA
TGCCAATGTTGTTTAAGAAACAGTTATGATCCTAAACTTTTTGGATAATCTTTTATATTTCTGACCTTTGAATTTAA
TCATTGTTCTTAGATTAAAATAAAATATGCTATTGAAACTA

<210> SEQ ID NO 1263

<211> Length : 233

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1263
>Z44808_PEA_1_node_41
CACCTCATAATCATGTGAAAAAGACACTCAAAAACTACCATTTGAATGGATGGATGAAAATAACCTCCGTATATTCT
ACGAAGATGTTTAATAATAAATAGGTTTCGTTATAAGAGAATGTGTGTCACTTCGTCTCTTCCCTCACCCCCGAGAC
TTAGTGACAGTTATTTTTGACTTTTCCAACTATACTATTTGCCTAGAAAATGTGTCTATTAAATAGCGTATTGAGAA
AT

<210> SEQ ID NO 1264

<211> Length : 51

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1264
>Z44808_PEA_1_node_11
ATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGAAG

<210> SEQ ID NO 1265

<211> Length : 75

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1265

>Z44808_PEA_1_node_13
ATATTGCATCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAGAATTCAG

<210> SEQ ID NO 1266

<211> Length : 83

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1266
>Z44808_PEA_1_node_18
GTACGAGCAGCCGAAATGTGACAACACGGCCAGGGCCCACCCAGCCAAAGCCCGGGACCTGTACAAGGGCCGCCAGC
TACAAG

<210> SEQ ID NO 1267

<211> Length : 103

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1267
>Z44808_PEA_1_node_22
GTTGTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGTTCTGGACGCGCTGTCCACGGACATGGTCCACGCCGCC
TCCGACCCCTCCTCCTCGTCAGGCAG

<210> SEQ ID NO 1268

<211> Length : 95

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1268
>Z44808_PEA_1_node_26
GAAGTAAGAGAAACCTGTGATGGCCAGAGCCCAGATGTTCTTAGGAGGCAAGCCAGGAGAAGCCGGGTCTGACTTTT
CAGCTCAGAGACAGCACT

<210> SEQ ID NO 1269
<211> Length : 38
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1269
>Z44808_PEA_1_node_30
CCCCCAGAGGTCATGCTGAAAGTACGTCTAATAGACAG

<210> SEQ ID NO 1270
<211> Length : 75
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1270
>Z44808_PEA_1_node_34
CTGATCCTCCTACCTGGTCCACCCCAGGGCTACCGGAAGGTAAAATCTTCACCTGAACCAATTATGAGCAGTCTC

<210> SEQ ID NO 1271
<211> Length : 19
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1271
>Z44808_PEA_1_node_35
CTTACTGAAGGTACAGCCG

<210> SEQ ID NO 1272

<211> Length : 65

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1272
>Z44808_PEA_1_node_39
CTACTGTGGTTGAGAGGAAAGGTGTCTTTTTATTGCTTCTAGAGACGTTGAAAGTGTGACCTGAG

<210> SEQ ID NO 1273

<211> Length : 107

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1273
>Z44808_PEA_1_node_4
ACGTGTCCAGGTGTGTGGCCGAAAGGAAGTATACCCAGGAGCAAGCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCT
GAGTGCAATGACGACGGCACCTACAGTCAG

<210> SEQ ID NO 1274

<211> Length : 100

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1274

>Z44808_PEA_1_node_6
GTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAGGCCCATCAGCGGCACTGCCGTGGC
CCACAAGACGCCCCGGTGCCCGG

<210> SEQ ID NO 1275

<211> Length : 48

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1275
>Z44808_PEA_1_node_8
GTTCCGTAAATGAAAAGTTACCCCAACGCGAAGGCACAGGAAAAACAG

Segment nucleic acid sequences:

<210> SEQ ID NO 1276

<211> Length : 126

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1276
>HSU33147_PEA_1_node_0
GTGCTCACCTCCACAGCGGCTTCCTTGATCCTTGCCACCCGCGACTGAACACCGACAGCAGCAGCCTCACCATGAAG
TTGCTGATGGTCCTCATGCTGGCGGCCCTCTCCCAGCACTGCTACGCAG

<210> SEQ ID NO 1277

<211> Length : 179

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1277

>HSU33147_PEA_1_node_2
GCTCTGGCTGCCCCTTATTGGAGAATGTGATTTCCAAGACAATCAATCCACAAGTGTCTAAGACTGAATACAAAGAA
CTTCTTCAAGAGTTCATAGACGACAATGCCACTACAAATGCCATAGATGAATTGAAGGAATGTTTTCTTAACCAAAC
GGATGAAACTCTGAGCAATGTTGAG

<210> SEQ ID NO 1278

<211> Length : 593

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1278
>HSU33147_PEA_1_node_4
GTAATTTCATTTTCTTCCTATAAGCTTTTTAAATCCCCTGACCAGGGACAAGTGGGCTCTTCATTTCTCACTGACAA
TGCCAAAGCCACTAGTGAACAAGCCTTTTCTTACATTGGTTAATTTAGTTGAATGGTTAGTCTAATGACTTTGCCAT
CAAGAAAAACATCCAGTGTCCCTGTGTTGTCACTCTACCCAGAGAATCCTCAGTGGATGATAAATGAATAGGGCAAG
AGAGGAAAAGGAAAGGTCGGTAGAAGTCTTACCTATCCCCAGAGCTCTCTAATTCATGCTCACAAACACAGACACAA
TCACACAAACACAGAAACACACATACACACATCCAGACACATGCAAACACACAGACACAGTCACAATCACACAAACA
CACACACATTCAGACATACACAAACATAGACAGACAGGCAAAGACACAGACACAGACACAGACACAATCACACCAGC
ACACAATCATCCAGACACAAACACAAACACACAGACAGAACCACACAACCACAGAAACACAGAGACACACACAAACA
CACTCAGACACACACATACAAACATATGTTCACTCTCTACAGAAAAAACAATTT

<210> SEQ ID NO 1279

<211> Length : 211

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1279

>HSU33147_PEA_1_node_7

CAATTAATATATGACAGCAGTCTTTGTGATTTATTTTAACTTTCTGCAAGACCTTTGGCTCACAGAACTGCAGGGTA
TGGTGAGAAACCAACTACGGATTGCTGCAAACCACACCTTCTCTTTCTTATGTCTTTTTACTACAAACTACAAGACA
ATTGTTGAAACCTGCTATACATGTTTATTTTAATAAATTGATGGCAAAAAAAAAAAT

<210> SEQ ID NO 1280

<211> Length : 9

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1280

>HSU33147_PEA_1_node_3

GTGTTTATG

<210> SEQ ID NO 1281
<211> Length : 152
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1281
>H61775_P16
MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWLRFGFLLPIFIQFGLYSPRI
DPDYVGDCGFPAFRELKRAETVSPVFFTRRCIWEDLKSTGFSPAGGGRPPGGGPRTQEDSGLPCWRSSCSVTLQV

<210> SEQ ID NO 1282
<211> Length : 83
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1282
>H61775_P17
MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWLRFGFLLPIFIQFGLYSPRI
DPDYVG

<210> SEQ ID NO 1283
<211> Length : 496
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1283
>M85491_PEA_1_P13
MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIRTYQVCNVFESSQNNWLRTKF
IRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGG
RVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAE
EVDVPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPINSRTTSEGATNCVCRNG
YYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDSGGREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQ
LGLTEPRIYISDLLAHTQYTFEIQAVNGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVDSITLSWSQPDQP
NGVILDYELQYYEKVPIGWVLSPSPTSLRAPLPG

<210> SEQ ID NO 1284
<211> Length : 301
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1284
>M85491_PEA_1_P14
MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIRTYQVCNVFESSQNNWLRTKF
IRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGG
RVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAE
EVDVPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRERQDLTMLSRLVLNSWPQMILPPQPPKVLEL

<210> SEQ ID NO 1285
<211> Length : 283
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1285
>T39971_P6
MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAECKPQVTRGDVFTMPEDEYTV
YDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEE
LCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFE

DGVLDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGTQGVVGD

<210> SEQ ID NO 1286
<211> Length : 447
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1286

>T39971_P9
MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAECKPQVTRGDVFTMPEDEYTV
YDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEE
LCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFE
DGVLDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSAVFEHFAMM
QRDSWEDIFELLFWGRTSGMAPRPSLAKKQRFRHRNRKGYRSQRGHSRGRNQNSRRPSRATWLSLFSSEESNLGANN
YDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL

<210> SEQ ID NO 1287
<211> Length : 363
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1287
>T39971_P11
MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAECKPQVTRGDVFTMPEDEYTV
YDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEE
LCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFE
DGVLDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSAVFEHFAMM
QRDSWEDIFELLFWGRTSDKYYRVNLRTRRVDTVDPPYPRSIAQYWLGCPAPGHL

<210> SEQ ID NO 1288
<211> Length : 238
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1288
>T39971_P12
MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAECKPQVTRGDVFTMPEDEYTV

YDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEE
LCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKVPGAVGQG
RKHLGRV

<210> SEQ ID NO 1289
<211> Length : 790
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1289
>Z21368_PEA_1_P2
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSA
PQFSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELEN
TYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSV
LKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQ
CIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVH
TRQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTH
KCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQE
KLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNPHKYSAHGR
TRHFESATRTTNGAQKLSRI

<210> SEQ ID NO 1290
<211> Length : 791
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1290
>Z21368_PEA_1_P5
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELAFFGKYLNEYNGSYIPPGWR
EWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQF
SKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYI
IYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKL
LDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIE
DTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQ
TRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTHKCF
ILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLK

SHLHPFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNN
NTYWCLRTVNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDV
GNKDGGSYDLHRGQLWDGWEG

<210> SEQ ID NO 1291
<211> Length : 416
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1291
>Z21368_PEA_1_P15
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSA
PQFSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELEN
TYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSV
LKLLDPEKPGNRFRTNKKAKIWRDTFLVERG

<210> SEQ ID NO 1292
<211> Length : 410
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1292
>Z21368_PEA_1_P16
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSA
PQFSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELEN
TYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSV
LKLLDPEKPGNRCVIVPPLSQPQIH

<210> SEQ ID NO 1293
<211> Length : 210
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1293

>Z21368_PEA_1_P22
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKARYDGDQPRCAPRPRGLSPTVF

<210> SEQ ID NO 1294
<211> Length : 145
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1294
>Z21368_PEA_1_P23
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTGLLHRLNH

<210> SEQ ID NO 1295

SEQ ID NO: 1295
HPRT1 Forward primer
TGACACTGGCAAAACAATGCA

SEQ ID NO: 1296
HPRT1 Reverse primer
GGTCCTTTTCACCAGCAAGCT

SEQ ID NO: 1297
HPRT1-amplicon
TGACACTGGCAAAACAATGCAGACTTTGCTTTCCTTGGTCAGGCAGTATAATCCAAAGATGGTCAAGGTCGCAAGCT
TGCTGGTGAAAAGGACC

SEQ ID NO: 1298
RPL19 Forward primer
TGGCAAGAAGAAGGTCTGGTTAG

<210> SEQ ID NO 1299
<211> Length : 141
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1299
>HUMGRP5E_P4
MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSSVSERGSLKQQLREYIRW
EEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSEDSSNFKDVGSKGKGSQREGRNPQLNQQ

<210> SEQ ID NO 1300
<211> Length : 142
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1300
>HUMGRP5E_P5
MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSSVSERGSLKQQLREYIRW
EEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSEDSSNFKDVGSKGKDSLLQVLNVKEGTPS

<210> SEQ ID NO 1301
<211> Length : 201
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1301
>D56406_PEA_1_P2
MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAHVPSWKMTLLNVCSLVNNLNSPAEETGE
VHEEELVARRKLPTALDGFSLEAMLTIYQLHKICHSRAFQHWEARWLTPVIPALWEAETGGSRGQEMETIPANTLIQ
EDILDTGNDKNGKEEVIKRKIPYILKRQLYENKPRRPYILKRDSYYY

<210> SEQ ID NO 1302
<211> Length : 168
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1302
>D56406_PEA_1_P5
MMAGMKIQLVCMLLLAFSSWSLCSEEEMKALEADFLTNMHTSKISKAHVPSWKMTLLNVCSLVNNLNSPAEETGEVH
EEELVARRKLPTALDGFSLEAMLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYENK
PRRPYILKRDSYYY

<210> SEQ ID NO 1303
<211> Length : 95
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1303
>D56406_PEA_1_P6
MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKLIQEDILDTGNDKNGKEEVIKRKIPYILKRQL
YENKPRRPYILKRDSYYY

Variant protein amino acid sequences:

<210> SEQ ID NO 1304
<211> Length : 33
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1304
>F05068_PEA_1_P7
MKLVSVALMYLGSLAFLGADTARLDVASEFRKK

<210> SEQ ID NO 1305
<211> Length : 83
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1305
>F05068_PEA_1_P8
MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSSSYPTGLADVKAGPAQTLIRPQDMKGAS
RSPEDR

<210> SEQ ID NO 1306
<211> Length : 180
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1306
>H14624_P15

MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLCHGIEYQNMRLPNLLGHETMKEVLEQAG
AWIPLVMKQCHPDTKKFLCSLFAPVCLDDLDETIQPCHSLCVQVKDRCAPVMSAFGFPWPDMLECDRFPQDNDLCIP
LASSDHLLPATEEGKPSLLLPHSLLG

<210> SEQ ID NO 1307
<211> Length : 381
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1307
>H38804_PEA_1_P5
MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQMTGSNEFKLNQPPEDGISSV
KFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDA
PIRCVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQQRR
ESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFAT
GGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYEMDDTEHPEDGIFIRQVTDAETKPK

<210> SEQ ID NO 1308
<211> Length : 385
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1308
>H38804_PEA_1_P17
MGRVRTLAGECSAQAQAQSLLAVVLSAPPSGGTPSARLSVRSPSPRDPWGLWAPVLQMTGSNEFKLNQPPEDGISSV
KFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGAVLDCAFYDPTHAWSGGLDHQLKMHDLNTDQENLVGTHDA
PIRCVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGTAGRRVLVWDLRNMGYVQQRR
ESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENNIEQIYPVNAISFHNIHNTFAT
GGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYEMDDTEHPEDGIFIRQVTDAETKPKSPCT

<210> SEQ ID NO 1309
<211> Length : 81
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1309
>HSENA78_P2
MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCVCLQTTQGVHPKMISNLQVFAIGPQCSK

VEVV

<210> SEQ ID NO 1310
<211> Length : 340
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1310
>HUMODCA_P9
MKSLTATSSMKVLLPRTFWTRKLMKFLLLLVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHV
GSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEP
GRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKYYS
SSIWGPTCDGLDRIVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEV
EEQDASTLPVSCAWESGMKRHRAACASASINV

<210> SEQ ID NO 1311
<211> Length : 283
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1311
>R00299_P3
MAEKALLCPSSAGLGTWPWVLNSAWPVLPLAVDQGVDWRPRGPVSSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDL
ELNPIRSKIVRAFFDNRNLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRI
TLEEYRNVVEELLSGNPHIEKESARSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIETKMHVRFLNM
ETMALCH

<210> SEQ ID NO 1312
<211> Length : 80
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1312
>W60282_PEA_1_P14
MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPTPASHLAMRQH
HHH

<210> SEQ ID NO 1313

<211> Length : 123
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1313
>Z41644_PEA_1_P10
MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVIITTKSVSRYRGQEHCLHPKL
QSTKRFIKWYNAWNEKRRYAPPLLTFLPTRPSCGSQDGKGPPHQVI

<210> SEQ ID NO 1314
<211> Length : 464
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1314
>Z44808_PEA_1_P5
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAKCKDPQLE
IAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCP
GSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPK
NDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQGCPGAK
KHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKP
KKCVKKFVEYCDVNNDKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQDAMVVSSRPKATTHRKSRTLS
RR

<210> SEQ ID NO 1315
<211> Length : 434
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1315
>Z44808_PEA_1_P6
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAKCKDPQLE
IAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCP
GSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPK
NDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQGCPGAK
KHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKP
KKCVKKFVEYCDVNNDKSISVQELMGCLGVAKEDGKADTKKRHRSKRNL

<210> SEQ ID NO 1316
<211> Length : 454
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1316
>Z44808_PEA_1_P7
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAKCKDPQLE
IAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCP
GSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPK
NDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQGCPGAK
KHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKP
KKCVKKFVEYCDVNNDKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQLLWLRGKVSFYCF
<210> SEQ ID NO 1317
<211> Length : 429
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1317
>Z44808_PEA_1_P11
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAKCKDPQLE
IAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCP
GSVNEKLPQREGTGKTDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKP
VQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMV
HAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDK
SISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQPRKQG

<210> SEQ ID NO 1318
<211> Length : 314
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1318
>AA161187_P1
MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWA
LTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQ
PICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGGK

DACFGDSGGPLACNKNGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEWIQKLMAQSGMSQPDPSWPLLFFPLLWAL
PLLGPV

<210> SEQ ID NO 1319
<211> Length : 326
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1319
>AA161187_P6
HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIRGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSH
VCGVSLLSHRWALTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVK
LSAPVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFG
DMVCAGNAQGGKDACFGDSGGPLACNKNGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEWIQKLMAQSGMSQPDPS
WPLLFFPLLWALPLLGPV

<210> SEQ ID NO 1320
<211> Length : 213
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1320
>AA161187_P13
MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWA
LTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQ
PICLQASTFEFENRTDCWVTGWGYIKEDEGSSGRHHKQLYVQPPLPQVQFPQGHLWRHG

<210> SEQ ID NO 1321
<211> Length : 307
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1321
>AA161187_P14
MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWA
LTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQ
PICLQASTFEFENRTDCWVTGWGYIKEDEGCCLSPSHYRPHSTAISPHPPGSSGRHHKQLYVQPPLPQVQFPQGHLW
RHGLCWQCPRREGCLLRECPCHHSQPRKASCVPVPYLTLMPTPGGGDCCPTLQMQKRRLGCCQGEEEDVHPVYPAP

<210> SEQ ID NO 1322
<211> Length : 265
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1322
>AA161187_P18
HTREGTLGGQKRAFPDGVEGEKGRGRAWGAASRGSAVPLTIRGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSH
VCGVSLLSHRWALTAAHCFETDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLS
APVTYTKHIQPICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDM
VCAGNAQGGKDACFVSVPATTPSPGKHPVSLCLI

<210> SEQ ID NO 1323
<211> Length : 188
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1323
>AA161187_P19
MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWA
LTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQ
PICLQASTFEFENRTDCWVTGWGYIKEDEDKRTQ

<210> SEQ ID NO 1324
<211> Length : 354
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1324
>R66178_P3
MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANPLPSVKITQVTWQKSTNGSKQN
VAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQ
AVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYH
MDRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTLFFKGPINY
SLAGTYICEATNPIGTRSGQVEVNITGEGHSLPISPGVLQTQNCGP

<210> SEQ ID NO 1325

<211> Length : 352
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1325
>R66178_P4
MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANPLPSVKITQVTWQKSTNGSKQN
VAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQ
AVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYH
MDRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTLFFKGPINY
SLAGTYICEATNPIGTRSGQVEVNITAFCQLIYPGKGRTRARMF

<210> SEQ ID NO 1326
<211> Length : 363
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1326
>R66178_P8
MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANPLPSVKITQVTWQKSTNGSKQN
VAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQ
AVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYH
MDRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTLFFKGPINY
SLAGTYICEATNPIGTRSGQVENSPTPRLLPNMGGAPGRCPRPSLGAWRGASCWC

<210> SEQ ID NO 1327
<211> Length : 398
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1327
>HUMPHOSLIP_PEA_2_P10
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEKVYDFLSTFI
TSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMKDPVASTSNLDMDFRGAFFPLTERNWSLP
NRAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLDMLLRATYFGSIVLLSPAVIDSPLKLE
LRVLAPPRCTIKPSGTTISVTASVTIALVPPDQPEVQLSSMTMDARLSAKMALRGKALRTQLDLRRFRIYSNHSALE
SLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVTNHAGFLTIGADLHFAKGLREVIEKNRPADV
RASTAPTPSTAAV

<210> SEQ ID NO 1328
<211> Length : 432
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1328
>HUMPHOSLIP_PEA_2_P12
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEVKVTELQLTS
SELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMH
AAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMKDPVASTSNLD
MDFRGAFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLDMLLRATYF
GSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASVTIALVPPDQPEVQLSSMTMDARLSAKMALRGKALR
TQLDLRRFRIYSNHSALESLALIPLQAPLKTMLQIGVMPMLNGKAGV

<210> SEQ ID NO 1329
<211> Length : 52
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1329
>HUMPHOSLIP_PEA_2_P30
MALFGALFLALLAGAHAEFPGRGCAFWSKSWRLSPFRTCGAKKATSTTTSLR

<210> SEQ ID NO 1330
<211> Length : 98
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1330
>HUMPHOSLIP_PEA_2_P31
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEPGLERGADKF
PVVGGSSLFLALDLTLRPPVG

<210> SEQ ID NO 1331
<211> Length : 200
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1331
>HUMPHOSLIP_PEA_2_P33
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEVKVTELQLTS
SELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMH
AAFGGTFKKVYDFLSTFITSGMRFLLNQQVWAATGRRVARVGMLSL

<210> SEQ ID NO 1332
<211> Length : 217
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1332
>HUMPHOSLIP_PEA_2_P34
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEVKVTELQLTS
SELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMH
AAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVLWTSLLALTIPS

<210> SEQ ID NO 1333
<211> Length : 148
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1333

>HUMPHOSLIP_PEA_2_P35
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEVKVTELQLTS
SELDFQPQQELMLQITNASLGLRFRRQLLYWFLKVYDFLSTFITSGMRFLLNQQVWAATGRRVARVGMLSL

<210> SEQ ID NO 1334
<211> Length : 258
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1334
>AI076020_P1

MLLVLVVLIPVLVSSGGPEGHYEMLGTCRMVCDPYPARGPGAGARTDGGDALSEQSGAPPPSTLVQGPQGKPGRTGK
PGPPGPPGDPGPPGPVGPPGEKGEPGKPGPPGLPGAGGSGAISTATYTTVPRVAFYAGLKNPHEGYEVLKFDDVVTN
LGNNYDAASGKFTCNIPGTYFFTYHVLMRGGDGTSMWADLCKNGQVRASAIAQDADQNYDYASNSVILHLDAGDEVF
IKLDGGKAHGGNSNKYSTFSGFIIYSD

<210> SEQ ID NO 1335
<211> Length : 140
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1335
>T23580_P5
MTLFPLTFLVQESPAESERLFRGAASPGTERNRPADGQQQGSLPGGHDRVRDAQADHVGRGILPLVERTNVPHHGLY
EKEIVSHLLTFSSFSKPSVPGFCKCCISAENPRCLLLPPPVHLELCKDSASVFLSSSGPRVSV

<210> SEQ ID NO 1336
<211> Length : 919
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1336
>M79217_PEA_1_P1
MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEADEAGKRIFGPRVGNELCEVK
HVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLS
LPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYVTEN
ADIACLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQSTFYTVQY
RPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKA
VQDSKLDQVLVEFTCKNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPV
VLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQ
IPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDFYRSWNCAPGPFHLFPH
TPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEFQAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVV
VVWNSPKLPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARDRIVG
FPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKP
PIKVTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI

210> SEQ ID NO 1337
<211> Length : 907
<212> Type : PRT

<213> Organism : Homo sapiens

<400> sequence : 1337
>M79217_PEA_1_P2
MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEADEAGKRIFGPRVGNELCEVK
HVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLS
LPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYVTEN
ADIACLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQSTFYTVQY
RPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKA
VQDSKLDQVLVEFTCKNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPV
VLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQ
IPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDFYRSWNCAPGPFHLFPH
TPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEFQAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVV
VVWNSPKLPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARDRIVG
FPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRC
PGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI

<210> SEQ ID NO 1338
<211> Length : 212
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1338
>M79217_PEA_1_P4
PELRQPARLGLPECWDYRHEPRCPAQMGSHFIVQAGLKLLASSKPPKCWDYRVWREARDRIVGFPGRYHAWDIPHQS
WLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKPPIKVTSRWTFRCPG
CPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI

<210> SEQ ID NO 1339
<211> Length : 812
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1339
>M79217_PEA_1_P8
MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEADEAGKRIFGPRVGNELCEVK
HVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLS
LPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYVTEN

ADIACLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQSTFYTVQY
RPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKA
VQDSKLDQVLVEFTCKNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPV
VLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQ
IPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDFYRSWNCAPGPFHLFPH
TPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEFQAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVV
VVWNSPKLPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARDRIVG
FPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKVRKSW

<210> SEQ ID NO 1340
<211> Length : 107
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1340
>M79217_PEA_1_P11
MGKRRHRPRVSLAVCGPPLATLSWLAGCRSASFSGWPACCSGLGCLTITPSQGSAGHCERWLPGQCSSVWTVPQARA
HQLGSCPEGWDLSGSCAGQSGELLVCGGKL

<210> SEQ ID NO 1341
<211> Length : 725
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1341
>M62096_PEA_1_P4
MATYIHVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSVF
NEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAKD
QKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRDY
EKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQELS
NHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSNR
KMNASERELAACQLLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRLQ
DAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLL
LLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVH
KQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKP
IRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK

<210> SEQ ID NO 1342

<211> Length : 674
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1342
>M62096_PEA_1_P5
MTRILQDSLGGNCRTTIVICCSPSVFNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVI
QHLEMELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQ
QSQLAEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQ
LTDELAQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYI
SKMKSEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEEL
AKLRAQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLE
QEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDND
DGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQ
EVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK

<210> SEQ ID NO 1343
<211> Length : 593
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1343
>M62096_PEA_1_P3
MELNRWRNGEAVPEDEQISAKDQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQL
AEKLKQQMLDQDELLASTRRDYEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDE
LAQKTTTLTTTQRELSQLQELSNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMK
SEVKSLVNRSKQLESAQMDSNRKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLR
AQEKMHEVSFQDKEKEHLTRLQDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKL
SSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGG
SAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDR
IKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK

<210> SEQ ID NO 1344
<211> Length : 239
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1344

>M62096_PEA_1_P7
MTQNFRLMWNILLFPLNFSLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL
QTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKVHKQLVRDNADLRCELPKLEKRLRAT
AERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAKPIRPGHYPASSPTAVHAIRGGGGSS
SNSTHYQK

<210> SEQ ID NO 1345
<211> Length : 737
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1345
>M62096_PEA_1_P8
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQEQVYNACAKQIVKDVLEG
YNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNL
AVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLV
DLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSV
FNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAK
DQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRD
YEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQEL
SNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSN
RKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRL
QDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRE

<210> SEQ ID NO 1346
<211> Length : 514
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1346
>M62096_PEA_1_P9
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQEQVYNACAKQIVKDVLEG
YNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNL
AVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLV
DLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSV
FNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAK
DQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDEVKNAIYFF
FHKVLLLLFVVDVCSRNLIGIEAFHNYRIMWKFLGRCPFTASYKLIITEFRK

<210> SEQ ID NO 1347
<211> Length : 125
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1347
>M62096_PEA_1_P10
MTQNFRLMWNILLFPLNFSLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKLLLLNDKREQAREDLKGLEETVSREL
QTLHNLRKLFVQDLTTRVKKVSSLCLNGTEKKIKDGREESFSVEISLA

<210> SEQ ID NO 1348
<211> Length : 385
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1348
>M62096_PEA_1_P11
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQEQVYNACAKQIVKDVLEG
YNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNL
AVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLV
DLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSV
FNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNDFLAAHVFGKLLE

<210> SEQ ID NO 1349
<211> Length : 324
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1349
>M62096_PEA_1_P12
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQEQVYNACAKQIVKDVLEG
YNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNL
AVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLV
DLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSV
FNEAETKSTLMFGQRV

<210> SEQ ID NO 1350

<211> Length : 519
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1350
>M78076_PEA_1_P3
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDGE

<210> SEQ ID NO 1351
<211> Length : 541
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1351
>M78076_PEA_1_P4
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGECLTVNPSLQIPL
NP

<210> SEQ ID NO 1352
<211> Length : 544
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1352
>M78076_PEA_1_P12
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD

PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGECVCSKGFPFPLI
GDSEG

<210> SEQ ID NO 1353
<211> Length : 619
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1353
>M78076_PEA_1_P14
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEK
MNPLEQYERKVNASVPRGFPFHSSEIQRDELVRGGTAGYLGEETRGQRPGCDSQSHTGPSKKPSAPSPLPAGTSWDR
GVP

<210> SEQ ID NO 1354
<211> Length : 597
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1354
>M78076_PEA_1_P21
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVD
PEKAQQMRFQVHTHLQVIEERVNQSLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDS

KDDTPMTLPKGSTEQDAASPEKEKMNPLEQYERKVNASVPRGFPFHSSEIQRDELAPAGTGVSREAVSGLLIMGAGG
GSLIVLSMLLLRRKKPYGAISHGVVEVDPMLTLEEQQLRELQRHGYENPTYRFLEERP

<210> SEQ ID NO 1355
<211> Length : 498
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1355
>M78076_PEA_1_P24
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIRECLLPWLPLQISEGRS

<210> SEQ ID NO 1356
<211> Length : 588
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1356
>M78076_PEA_1_P2
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVLTSFQLPNAPLFL
RRPRLRLFSCPLDPLSVSWTPSYPLNTASLPLPSLSAQLPDPETWTLTCCVFDPCFLALGFLLPPPSILCSVPWIFT
AFPRIVFFFFFFLRQVLALSPRQESSVRSWLIATSTSWVQAILLPQPLE

<210> SEQ ID NO 1357
<211> Length : 505
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1357
>M78076_PEA_1_P25
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD
PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQPQNPNSQPRAAGSL
EVIISHPFVRRLEILISPFQFQNSIPKNSQIVPAASPRGTSSP

<210> SEQ ID NO 1358
<211> Length : 62
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1358
>T99080_PEA_4_P1
MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQLFTI

<210> SEQ ID NO 1359
<211> Length : 64
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1359
>T99080_PEA_4_P2
SGRGGLRALVSRWRGGPGVILAAGGEDKEGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQLFTI

<210> SEQ ID NO 1360
<211> Length : 129
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1360
>T99080_PEA_4_P5
MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGT
VQGQLQGPISKVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK

<210> SEQ ID NO 1361
<211> Length : 73
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1361
>T99080_PEA_4_P8
MQAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHIDKANFNNEKVILKLDYSDFQIVK

<210> SEQ ID NO 1362
<211> Length : 87
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1362
>T99080_PEA_4_P9
SGRGGLRALVSRWRGGPGVILAAGGEDKEGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQEMTVENRIAETHSKSCV
PVSFATSYAG

<210> SEQ ID NO 1363
<211> Length : 80
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1363
>T99080_PEA_4_P10
SGRGGLRALVSRWRGGPGVILAAGGEDKEGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQLCFWEAEEGGSFEPGRL
RLQ

<210> SEQ ID NO 1364
<211> Length : 94
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1364
>T99080_PEA_4_P12
SGRGGLRALVSRWRGGPGVILAAGGEDKEGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQEMTVENRIAETHSKSCV

PVSFATSYAGINEFKRI

<210> SEQ ID NO 1365
<211> Length : 69
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1365
>T99080_PEA_4_P13
SGRGGLRALVSRWRGGPGVILAAGGEDKEGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQVCGLQALGW

<210> SEQ ID NO 1366
<211> Length : 85
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1366
>T99080_PEA_4_P14
MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQEMTVENRIAETHSKSCVPV
SFATSYAG

<210> SEQ ID NO 1367
<211> Length : 78
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1367
>T99080_PEA_4_P15
MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQLCFWEAEEGGSFEPGRLRL
Q

<210> SEQ ID NO 1368
<211> Length : 92
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1368
>T99080_PEA_4_P16

MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQEMTVENRIAETHSKSCVPV
SFATSYAGINEFKRI

<210> SEQ ID NO 1369
<211> Length : 67
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1369
>T99080_PEA_4_P17
MPASARLAGAGLLLAFLRALGCAGRAPGLSMAEGNTLISVDYEIFGKVQGVFFRKHTQVCGLQALGW

<210> SEQ ID NO 1370
<211> Length : 1305
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1370
>T08446_PEA_1_P18
MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRGPFPRLADCAHFHYENVDF
GHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSYDDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ
MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIV
SVIDMPPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLAGLLR
TFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVVDGIYRLSGVSSNIQRLRHEFDS
ERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLL
RHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDP
AGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALG
RGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSC
ESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTP
GDPAPPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPG
RSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALALAE
RAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQ
RPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSF
QPSSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLS
YPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPW
GPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC

<210> SEQ ID NO 1371
<211> Length : 246
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1371
>T08446_PEA_1_P19
MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRGPFPRLADCAHFHYENVDF
GHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSYDDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ
MLVPLLLQYLETLSGLVDSNLNCGPVLTWMEVGLGRGLGDSEWVRGCVCHHAQHREILDGNRVASAVEDEGAEVDGE
AFRWGSLWVGESWDM

<210> SEQ ID NO 1372
<211> Length : 1081
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1372
>HUMCA1XIA_P14
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEANIVDDFQE
YNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYE
YKEYEDKPTSPPNEEFGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPT
GPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGR
PGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDRGFDGLPGLPGDKGHRG
ERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGPRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQ
GNPGPQGLPGPQGPIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPPGPQGPIGYPGPRGVKGADGV
RGLKGSKGEKGEDGFPGFKGDMGLKGDRGEVGQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLPGYPG
RQGPKGSTGFPGFPGANGEKGARGVAGKPGPRGQRGPTGPRGSRGARGPTGKPGPKGTSGGDGPPGPPGERGPQGPQ
GPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGA
AGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQGPPGPVVSMMIINSQTIMVVNYSSSFIT
LML

<210> SEQ ID NO 1373
<211> Length : 729
<212> Type : PRT

<213> Organism : Homo sapiens

<400> sequence : 1373
>HUMCA1XIA_P15
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEANIVDDFQE
YNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYE
YKEYEDKPTSPPNEEFGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPT
GPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGR
PGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDRGFDGLPGLPGDKGHRG
ERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGPRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQ
GNPGPQGLPGPQGPIGPPGEKMCCNLSFGILIPLQK

<210> SEQ ID NO 1374
<211> Length : 738
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1374
>HUMCA1XIA_P16
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEANIVDDFQE
YNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYE
YKEYEDKPTSPPNEEFGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPT
GPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGR
PGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDRGFDGLPGLPGDKGHRG
ERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGMAGVDGPPGPKGNMGPQGEPGPPGQQGNPGPQGLPGPQGPIGPP
GEKVSFSFSLFYKKVIKFACDKRFVGRHDERKVVKLSLPLYLIYE

<210> SEQ ID NO 1375
<211> Length : 273
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1375
>HUMCA1XIA_P17
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEVRSTRPEKVFVFQ

<210> SEQ ID NO 1376
<211> Length : 154
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1376
>T11628_PEA_1_P2
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1377
<211> Length : 99
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1377
>T11628_PEA_1_P5
MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAM
NKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1378
<211> Length : 135
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1378
>T11628_PEA_1_P7
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKG

<210> SEQ ID NO 1379

<211> Length : 154
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1379
>T11628_PEA_1_P10
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1380
<211> Length : 315
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1380
>HUMCEA_PEA_1_P4
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVIL
NVLCEYICSSLAQAASPNPQGQRQDFSVPLRFKYTDPQPWTSRLSVTFCPRKTWADQVLTKNRRGGAASVLGGSGST
PYDGRNR

<210> SEQ ID NO 1381
<211> Length : 719
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1381
>HUMCEA_PEA_1_P5
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVIL
NVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHNSDTGL
NRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTR
NDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQEL
FISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ
SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGANLNLSCH
SASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSGKWLPGASASYSGVESIW

FSPKSQEDIFFPSLCSMGTRKSQILS

<210> SEQ ID NO 1382
<211> Length : 569
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1382
>HUMCEA_PEA_1_P14
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVIL
NVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHNSDTGL
NRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTR
NDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQEL
FISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNNSNPVEDKDAVAFTCEPEVQNTTYLWWVNGQ
SLPVSPRLQLSNGNMTLTLLSCQKERCRIL

<210> SEQ ID NO 1383
<211> Length : 346
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1383
>HUMCEA_PEA_1_P19
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVIL
NVLYGPDTPIISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGR
NNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVALI

<210> SEQ ID NO 1384
<211> Length : 346
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1384
>HUMCEA_PEA_1_P20

MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTL
DVLYGPDTPIISPPDSSYLSGANLNLSCHSASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGR
NNSIVKSITVSASGTSPGLSAGATVGIMIGVLVGVALI

<210> SEQ ID NO 1385
<211> Length : 385
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1385
>R35137_PEA_1_PEA_1_PEA_1_P9
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDH
CRPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVRGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAV
PLIQEGAHGDGAALRRAAGACLLPLHLQGLHGRVRAYEAGGGSRAMARPSSPDGPPPPPHLTWPCAGAGSAAAMWRW

<210> SEQ ID NO 1386
<211> Length : 346
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1386
>R35137_PEA_1_PEA_1_PEA_1_P8
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDH
CRPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPYAGQQELAS
FHSTSKGYMGECVRTRRVGARGPWPGPPRPMGHPLLRT

<210> SEQ ID NO 1387
<211> Length : 271
<212> Type : PRT

<213> Organism : Homo sapiens

<400> sequence : 1387
>R35137_PEA_1_PEA_1_PEA_1_P11
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARSG
FGQREGTYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS

<210> SEQ ID NO 1388
<211> Length : 399
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1388
>R35137_PEA_1_PEA_1_PEA_1_P2
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDH
CRPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVRGAGEREAGQQSAPVTPCALPGVPGQRVRRGFAV
PLIQEGAHGDGAALRRAAGACLLPLHLQGLHGRVRVPRRLCGGGEHGRCSAAADAEADECAAVPAGARTGPAGPGGQ
PARAHRPLLCAVPG

<210> SEQ ID NO 1389
<211> Length : 555
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1389
>R35137_PEA_1_PEA_1_PEA_1_P4
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDH
CRPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPYAGQQELAS
FHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCPPVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAE
LAAKAKLTEQVFNEAPGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGICVVPGSGFGQREG
TYHFRMTILPPLEKLRLLLEKLSRFHAKFTLESPGRLWSPLYLLLMPGGVGWGGCWAPASLQVPNKAVWQSDSKKEA
LAAAWPAPTCLPFLQA

<210> SEQ ID NO 1390
<211> Length : 139
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1390
>Z25299_PEA_2_P2
MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCPGKKRCCPDTCGIKCLDPVD
TPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLKCCMGMCGKSCVSPVKGKQGMRAH

<210> SEQ ID NO 1391
<211> Length : 156
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1391
>Z25299_PEA_2_P3
MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCPGKKRCCPDTCGIKCLDPVD
TPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLKCCMGMCGKSCVSPVKGEKRHHKQLRDQEVDPLEMRRHS
AG
<210> SEQ ID NO 1392
<211> Length : 89
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1392

>Z25299_PEA_2_P7
MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCPGKKRCCPDTCGIKCLDPVD
TPNPRGSLGSAQ
<210> SEQ ID NO 1393
<211> Length : 82
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1393
>Z25299_PEA_2_P10

MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCPGKKRCCPDTCGIKCLDPVD
TPNPT

<210> SEQ ID NO 1394
<211> Length : 496
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1394
>HSSTROL3_P4
MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRAD
IMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKA
LMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDAVSTIRGEL
FFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDAQGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRF
PVHAALVWGPEKNKIYFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADGALGVRQLVGGGHSSRFS
HLVVAGLPHACHRKSGSSSQVLCPEPSALLSVAG

<210> SEQ ID NO 1395
<211> Length : 382
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1395
>HSSTROL3_P5
MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRAD
IMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKA
LMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDAVSTIRGEL
FFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDAQGHIWFFQELGFPSSTGRDESLEHCRCQGLHK

<210> SEQ ID NO 1396
<211> Length : 370
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1396
>HSSTROL3_P7

MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRAD
IMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKA
LMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDAVSTIRGEL
FFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDAQGHIWFFQGTTGVSTPAPGV

<210> SEQ ID NO 1397
<211> Length : 301
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1397
>HSSTROL3_P8
MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRAD
IMIDFARYWHGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKA
LMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEVRPCLPVPLLLCWPL

<210> SEQ ID NO 1398
<211> Length : 354
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1398
>HSSTROL3_P9
MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRADIMIDFARYWHGDDLPF
DGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKALMSAFYTFRYPLSLSP
DDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDAVSTIRGELFFFKAGFVWRLRGGQL
QPGYPALASRHWQGLPSPVDAAFEDAQGHIWFFQGTTGVSTPAPGV

<210> SEQ ID NO 1399
<211> Length : 137
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1399
>HUMTREFAC_PEA_2_P7

MAARALCMLGLVLALLSSSSAEEYVGLSQQGLWQLTGLCLGQLQTSVPCQPRTGWTAATPMSPPRSATTGAAALTPG
SLECLGVSSPCRKQNAPSEAPPAAPGRGMRGSEHPCPAVIAARHCSSQLFCPFAPGKRFC

<210> SEQ ID NO 1400
<211> Length : 41
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1400
>HUMTREFAC_PEA_2_P8
MAARALCMLGLVLALLSSSSAEEYVGLWKVHLPKGEGFSSG

<210> SEQ ID NO 1401
<211> Length : 159
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1401
>HSS100PCB_P3
MHTVVSWSDDMFLPILIVFLSTWTHSSLAQGWVISTPCCCCSDLHPGPAWSHSENALLSHGQALRALPAWPSQRRPE
SLGPKQRGRTGLRVLAPCSLHTWGSTLAFPGSLCDFRAGPLGPLGLIIFICNRAMPLPCLVVLRSSCLCKQLVQCQH
EMGGP

<210> SEQ ID NO 1402
<211> Length : 187
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1402
>R20779_P2
MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNTAEIQHCLVNAGDVGCGVFECFENNSCE
IRGLHGICMTFLHNAGKFDAQGKSFIKDALKCKAHALRHRFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQENTR
VIVEMIHFKDLLLHECYKIEITMPKRRKVKLRD

<210> SEQ ID NO 1403
<211> Length : 449
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1403
>R38144_PEA_2_P6
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVS
MPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATG
DPTLLELGRDAVESIEKISKVECGFATLASFSHMSDQRSARPQAGQPHGVVLPGRDCEIPLPPV

<210> SEQ ID NO 1404
<211> Length : 341
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1404
>R38144_PEA_2_P13
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVS
MPVFQSLEAYWPGLQNLLKAQCTSTVPRGIPPS

<210> SEQ ID NO 1405
<211> Length : 287
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1405
>R38144_PEA_2_P15
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEPHWRH
<210> SEQ ID NO 1406
<211> Length : 433
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1406
>R38144_PEA_2_P19
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVS
MPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATG
DPTLLELGRDAVESIEKISKVECGFATKRSRSVAQAGVQWCDHDSPQP

<210> SEQ ID NO 1407
<211> Length : 418
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1407
>R38144_PEA_2_P24
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIREYNKAIRNYTRFDDWYLWVQMYKGTVSMPVFQS
LEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLE
LGRDAVESIEKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNNGSTFDAVITPYGECILGA
GGYIFNTEAHPIDPAALHCCQRLKEEQWEVEDLMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPENHDQARE
RKPAKQKVPLLSCPSQPFTSKLALLGQVFLDSS

<210> SEQ ID NO 1408
<211> Length : 60
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1408
>R38144_PEA_2_P36
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRFWGMSQNSKEWLKCSRTAWTLILM

<210> SEQ ID NO 1409
<211> Length : 112
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1409
>R11723_PEA_1_P2
MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPN
LVGHPAYGQCHNNQPWADTSRRERQRKEKHSMRTQ
<210> SEQ ID NO 1410
<211> Length : 222
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1410
>R11723_PEA_1_P6
MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSAGIMYRKSCASSAAC
LIASAGSPCRGLAPGREEQRALHKAGAVGGGVRMYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEV
EKRLREGEEDHVRPEVGPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRERQRKEKHSMRTQ

<210> SEQ ID NO 1411
<211> Length : 93
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1411
>R11723_PEA_1_P7
MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSAGSHCVTRLECSGTI
SAHCNLCLPGSNDHPT

<210> SEQ ID NO 1412
<211> Length : 84
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1412
>R11723_PEA_1_P13
MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSADTKRTNTLLFEMRH
FAKQLTT

<210> SEQ ID NO 1413
<211> Length : 90
<212> Type : PRT

<213> Organism : Homo sapiens

<400> sequence : 1413
>R11723_PEA_1_P10
MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVMEQSADRVSLCHEAGVQWN
NFSTLQPLPPRLK

<210> SEQ ID NO 1414
<211> Length : 111
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1414
>R16276_PEA_1_P7
MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPGQCPATPPTCAPGVRAVLDGCSCCLVCARQRGESCSD
LEPCDESSGLYCDRSADPSNQTGICTGNPAPSAV

<210> SEQ ID NO 1415
<211> Length : 111
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1415
>HSU33147_PEA_1_P5
MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAIDELKECFLNQTDETLSNV
EQLIYDSSLCDLF

<210> SEQ ID NO 1416
<211> Length : 93
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1416
>Mammaglobin A precursor
MKLLMVLMLAALSQHCYAGSGCPLLENVISKTINPQVSKTEYKELLQEFIDDNATTNAIDELKECFLNQTDETLSNV
EVFMQLIYDSSLCDLF

<210> SEQ ID NO 1417
<211> Length : 1055
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1417

>Ephrin type-B receptor 2 [precursor]
MALRRLGAALLLLPLLAAVEETLMDSTTATAELGWMVHPPSGWEEVSGYDENMNTIRTYQVCNVFESSQNNWLRTKF
IRRRGAHRIHVEMKFSVRDCSSIPSVPGSCKETFNLYYYEADFDSATKTFPNWMENPWVKVDTIAADESFSQVDLGG
RVMKINTEVRSFGPVSRSGFYLAFQDYGGCMSLIAVRVFYRKCPRIIQNGAIFQETLSGAESTSLVAARGSCIANAE
EVDVPIKLYCNGDGEWLVPIGRCMCKAGFEAVENGTVCRGCPSGTFKANQGDEACTHCPINSRTTSEGATNCVCRNG
YYRADLDPLDMPCTTIPSAPQAVISSVNETSLMLEWTPPRDSGGREDLVYNIICKSCGSGRGACTRCGDNVQYAPRQ
LGLTEPRIYISDLLAHTQYTFEIQAVNGVTDQSPFSPQFASVNITTNQAAPSAVSIMHQVSRTVDSITLSWSQPDQP
NGVILDYELQYYEKELSEYNATAIKSPTNTVTVQGLKAGAIYVFQVRARTVAGYGRYSGKMYFQTMTEAEYQTSIQE
KLPLIIGSSAAGLVFLIAVVVIAIVCNRRGFERADSEYTDKLQHYTSGHMTPGMKIYIDPFTYEDPNEAVREFAKEI
DISCVKIEQVIGAGEFGEVCSGHLKLPGKREIFVAIKTLKSGYTEKQRRDFLSEASIMGQFDHPNVIHLEGVVTKST
PVMIITEFMENGSLDSFLRQNDGQFTVIQLVGMLRGIAAGMKYLADMNYVHRDLAARNILVNSNLVCKVSDFGLSRF
LEDDTSDPTYTSALGGKIPIRWTAPEAIQYRKFTSASDVWSYGIVMWEVMSYGERPYWDMTNQDVINAIEQDYRLPP
PMDCPSALHQLMLDCWQKDRNHRPKFGQIVNTLDKMIRNPNSLKAMAPLSSGINLPLLDRTIPDYTSFNTVDEWLEA
IKMGQYKESFANAGFTSFDVVSQMMMEDILRLGVTLAGHQKKILNSIQVMRAQMNQIQSVEGQPLARRPRATGRTKR
CQPRDVTKKTCNSNDGKKKGMGKKKTDPGRGREIQGIFFKEDSHKESNDCSCGG

<210> SEQ ID NO 1418
<211> Length : 478
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1418

>Vitronectin precursor
MAPLRPLLILALLAWVALADQESCKGRCTEGFNVDKKCQCDELCSYYQSCCTDYTAECKPQVTRGDVFTMPEDEYTV
YDDGEEKNNATVHEQVGGPSLTSDLQAQSKGNPEQTPVLKPEEEAPAPEVGASKPEGIDSRPETLHPGRPQPPAEEE
LCSGKPFDAFTDLKNGSLFAFRGQYCYELDEKAVRPGYPKLIRDVWGIEGPIDAAFTRINCQGKTYLFKGSQYWRFE

DGVLDPDYPRNISDGFDGIPDNVDAALALPAHSYSGRERVYFFKGKQYWEYQFQHQPSQEECEGSSLSAVFEHFAMM
QRDSWEDIFELLFWGRTSAGTRQPQFISRDWHGVPGQVDAAMAGRIYISGMAPRPSLAKKQRFRHRNRKGYRSQRGH
SRGRNQNSRRPSRATWLSLFSSEESNLGANNYDDYRMDWLVPATCEPIQSVFFFSGDKYYRVNLRTRRVDTVDPPYP
RSIAQYWLGCPAPGHL

<210> SEQ ID NO 1419
<211> Length : 871
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1419

>Extracellular sulfatase Sulf-1 precursor
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSA
PQFSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELEN
TYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSV
LKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQ
CIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVH
TRQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTH
KCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQE
KLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTS
SNNNTYWCLRTVNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKN
LDVGNKDGGSYDLHRGQLWDGWEG

SEQ ID NO: 1420
RPL19 Reverse primer
TGATCAGCCCATCTTTGATGAG

<210> SEQ ID NO 1421

<211> Length : 148
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1421

>Gastrin-releasing peptide precursor
MRGSELPLVLLALVLCLAPRGRAVPLPAGGGTVLTKMYPRGNHWAVGHLMGKKSTGESSSVSERGSLKQQLREYIRW
EEAARNLLGLIEAKENRNHQPPQPKALGNQQPSWDSEDSSNFKDVGSKGKVGRLSAPGSQREGRNPQLNQQ

<210> SEQ ID NO 1422
<211> Length : 170
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1422

>Neurotensin/neuromedin N precursor [Contains: Large neuromedin N (NmN- 125); Neuromedin N (NmN) (NN); Neurotensin (NT); Tail peptide]
MMAGMKIQLVCMLLLAFSSWSLCSDSEEEMKALEADFLTNMHTSKISKAHVPSWKMTLLNVCSLVNNLNSPAEETGE
VHEEELVARRKLPTALDGFSLEAMLTIYQLHKICHSRAFQHWELIQEDILDTGNDKNGKEEVIKRKIPYILKRQLYE
NKPRRPYILKRDSYYY

<210> SEQ ID NO 1423
<211> Length : 185
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1423

>ADM precursor [Contains: Adrenomedullin (AM); Proadrenomedullin N-20 terminal peptide (ProAM-N20) (ProAM N-terminal 20 peptide) (PAMP)]
MKLVSVALMYLGSLAFLGADTARLDVASEFRKKWNKWALSRGKRELRMSSSYPTGLADVKAGPAQTLIRPQDMKGAS
RSPEDSSPDAARIRVKRYRQSMNNFQGLRSFGCRFGTCTVQKLAHQIYQFTDKDKDNVAPRSKISPQGYGRRRRRSL
PEAGPGRTLVSSKPQAHGAPAPPSGSAPHFL

<210> SEQ ID NO 1424
<211> Length : 328
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1424

>Mitotic checkpoint protein BUB3
MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGAVLDCAFYDPTHAWSGGLDH
QLKMHDLNTDQENLVGTHDAPIRCVEYCPEVNVMVTGSWDQTVKLWDPRTPCNAGTFSQPEKVYTLSVSGDRLIVGT
AGRRVLVWDLRNMGYVQQRRESSLKYQTRCIRAFPNKQGYVLSSIEGRVAVEYLDPSPEVQKKKYAFKCHRLKENNI
EQIYPVNAISFHNIHNTFATGGSDGFVNIWDPFNKKRLCQFHRYPTSIASLAFSNDGTTLAIASSYMYEMDDTEHPE
DGIFIRQVTDAETKPKSPCT

<210> SEQ ID NO 1425
<211> Length : 114
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1425

>Small inducible cytokine B5 precursor
MSLLSSRAARVPGPSSSLCALLVLLLLLTQPGPIASAGPAAAVLRELRCVCLQTTQGVHPKMISNLQVFAIGPQCSK
VEVVASLKNGKEICLDPEAPFLKKVIQKILDGGNKEN

<210> SEQ ID NO 1426
<211> Length : 461
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1426

>Ornithine decarboxylase
MNNFGNEEFDCHFLDEGFTAKDILDQKINEVSSSDDKDAFYVADLGDILKKHLRWLKALPRVTPFYAVKCNDSKAIV
KTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQVSQIKYAANNGVQMMTFDSEVELMKVARAHPKAKLVLR
IATDDSKAVCRLSVKFGATLRTSRLLLERAKELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYL
LDIGGGFPGSEDVKLKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDDEDE
SSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDRIVERCDLPEMHVGDWMLFEN
MGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDFPPEVEEQDASTLPVSCAWESGMKRHRAACASASINV

<210> SEQ ID NO 1427
<211> Length : 214
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1427

>Tescalcin
MGAAHSASEEVRELEGKTGFSSDQIEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFFDNRNLRKGPS
GLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSDGRITLEEYRNVVEELLSGNPHIEKESA
RSIADGAMMEAASVCMGQMEPDQVYEGITFEDFLKIWQGIDIETKMHVRFLNMETMALCH

<210> SEQ ID NO 1428
<211> Length : 250
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1428

>Kallikrein 11 precursor
MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTAAHCLKPRYIVHLGQHNL
QKEEGCEQTRTATESFPHPGFNNSLPNKDHRNDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCLISGWGSTSSPQ
LRLPHTLRCANITIIEHQKCENAYPGNITDTMVCASVQEGGKDSCQGDSGGPLVCNQSLQGIISWGQDPCAITRKPG
VYTKVCKYVDWIQETMKNN

<210> SEQ ID NO 1429
<211> Length : 99
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1429

>Small inducible cytokine B14 precursor
MRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPHCEEKMVIITTKSVSRYRGQEHCLHPKL
QSTKRFIKWYNAWNEKRRVYEE

<210> SEQ ID NO 1430
<211> Length : 446
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1430

>SPARC related modular calcium-binding protein 2 precursor
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAKCKDPQLE
IAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRPISGTAVAHKTPRCP
GSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNKNSVSSCDQEHQSALEEAKQPK
NDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPKCDNTARAHPAKARDLYKGRQLQGCPGAK
KHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEERVVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKP
KKCVKKFVEYCDVNNDKSISVQELMGCLGVAKEDGKADTKKRHTPRGHAESTSNRQPRKQG

<210> SEQ ID NO 1431
<211> Length : 314
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1431

>Testisin precursor

MGARGALLLALLLARAGLRKPESQEAAPLSGPCGRRVITSRIVGGEDAELGRWPWQGSLRLWDSHVCGVSLLSHRWA
LTAAHCFETYSDLSDPSGWMVQFGQLTSMPSFWSLQAYYTRYFVSNIYLSPRYLGNSPYDIALVKLSAPVTYTKHIQ
PICLQASTFEFENRTDCWVTGWGYIKEDEALPSPHTLQEVQVAIINNSMCNHLFLKYSFRKDIFGDMVCAGNAQGGK
DACFGDSGGPLACNKNGLWYQIGVVSWGVGCGRPNRPGVYTNISHHFEWIQKLMAQSGMSQPDPSWPLLFFPLLWAL
PLLGPV

<210> SEQ ID NO 1432
<211> Length : 517
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1432

>Poliovirus receptor related protein 1 precursor
MARMGLAGAAGRWWGLALGLTAFFLPGVHSQVVQVNDSMYGFIGTDVVLHCSFANPLPSVKITQVTWQKSTNGSKQN
VAIYNPSMGVSVLAPYRERVEFLRPSFTDGTIRLSRLELEDEGVYICEFATFPTGNRESQLNLTVMAKPTNWIEGTQ
AVLRAKKGQDDKVLVATCTSANGKPPSVVSWETRLKGEAEYQEIRNPNGTVTVISRYRLVPSREAHQQSLACIVNYH
MDRFKESLTLNVQYEPEVTIEGFDGNWYLQRMDVKLTCKADANPPATEYHWTTLNGSLPKGVEAQNRTLFFKGPINY
SLAGTYICEATNPIGTRSGQVEVNITEFPYTPSPPEHGRRAGPVPTAIIGGVAGSILLVLIVVGGIVVALRRRRHTF
KGDYSTKKHVYGNGYSKAGIPQHHPPMAQNLQYPDDSDDEKKAGPLGGSSYEEEEEEEGGGGGERKVGGPHPKYDE
DAKRPYFTVDEAEARQDGYGDRTLGYQYDPEQLDLAENMVSQNDGSFISKKEWYV

<210> SEQ ID NO 1433
<211> Length : 493
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1433

>Phospholipid transfer protein precursor
MALFGALFLALLAGAHAEFPGCKIRVTSKALELVKQEGLRFLEQELETITIPDLRGKEGHFYYNISEVKVTELQLTS
SELDFQPQQELMLQITNASLGLRFRRQLLYWFFYDGGYINASAEGVSIRTGLELSRDPAGRMKVSNVSCQASVSRMH
AAFGGTFKKVYDFLSTFITSGMRFLLNQQICPVLYHAGTVLLNSLLDTVPVRSSVDELVGIDYSLMKDPVASTSNLD
MDFRGAFFPLTERNWSLPNRAVEPQLQEEERMVYVAFSEFFFDSAMESYFRAGALQLLLVGDKVPHDLDMLLRATYF
GSIVLLSPAVIDSPLKLELRVLAPPRCTIKPSGTTISVTASVTIALVPPDQPEVQLSSMTMDARLSAKMALRGKALR

TQLDLRRFRIYSNHSALESLALIPLQAPLKTMLQIGVMPMLNERTWRGVQIPLPEGINFVHEVVTNHAGFLTIGADL
HFAKGLREVIEKNRPADVRASTAPTPSTAAV

<210> SEQ ID NO 1434
<211> Length : 258
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1434

>C1q-related factor precursor
MLLVLVVLIPVLVSGGPEGHYEMLGTCRMVCDPYPARGPGAGARTDGGDALSEQSGAPPPSTLVQGPQGKPGRTGK
PGPPGPPGDPGPPGPVGPPGEKGEPGKPGPPGLPGAGGSGAISTATYTTVPRVAFYAGLKNPHEGYEVLKFDDVVTN
LGNNYDAASGKFTCNIPGTYFFTYHVLMRGGDGTSMWADLCKNGQVRASAIAQDADQNYDYASNSVILHLDAGDEVF
IKLDGGKAHGGNSNKYSTFSGFIIYSD

<210> SEQ ID NO 1435
<211> Length : 199
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1435

>Neuronal protein NP25
MANRGPSYGLSREVQEKIEQKYDADLENKLVDWIILQCAEDIEHPPPGRAHFQKWLMDGTVLCKLINSLYPPGQEPI
PKISESKMAFKQMEQISQFLKAAETYGVRTTDIFQTVDLWEGKDMAAVQRTLMALGSVAVTKDDGCYRGEPSWFHRK
AQQNRRGFSEEQLRQGQNVIGLQMGSNKGASQAGMTGYGMPRQIM

<210> SEQ ID NO 1436
<211> Length : 919
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1436

>Exostosin-like 3
MTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEADEAGKRIFGPRVGNELCEVK
HVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSYKELMAQNQPKLS
LPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQATARANVYVTEN
ADIACLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRAMVAQSTFYTVQY
RPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESLRSSLQEARSFEEEMEGDPPADYDDRIIATLKA
VQDSKLDQVLVEFTCKNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATRLFEALEVGAVPV
VLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIFNTVLAMIRTRIQ
IPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDFYRSWNCAPGPFHLFPH
TPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEFQAALGGNVPREQFTVVMLTYEREEVLMNSLERLNGLPYLNKVV
VVWNSPKLPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGFRVWREARDRIVG
FPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIAMNFLVSHITRKP
PIKVTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDKTKCFKFI

<210> SEQ ID NO 1437
<211> Length : 931
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1437

>BAA25445
PDVIWGAGCRGLMTGYTMLRNGGAGNGGQTCMLRWSNRIRLTWLSFTLFVILVFFPLIAHYYLTTLDEADEAGKRIF
GPRVGNELCEVKHVLDLCRIRESVSEELLQLEAKRQELNSEIAKLNLKIEACKKSIENAKQDLLQLKNVISQTEHSY
KELMAQNQPKLSLPIRLLPEKDDAGLPPPKATRGCRLHNCFDYSRCPLTSGFPVYVYDSDQFVFGSYLDPLVKQAFQ
ATARANVYVTENADIACLYVILVGEMQEPVVLRPAELEKQLYSLPHWRTDGHNHVIINLSRKSDTQNLLYNVSTGRA
MVAQSTFYTVQYRPGFDLVVSPLVHAMSEPNFMEIPPQVPVKRKYLFTFQGEKIESLRSSLQEARSFEEEMEGDPPA
DYDDRIIATLKAVQDSKLDQVLVEFTCKNQPKPSLPTEWALCGEREDRLELLKLSTFALIITPGDPRLVISSGCATR
LFEALEVGAVPVVLGEQVQLPYQDMLQWNEAALVVPKPRVTEVHFLLRSLSDSDLLAMRRQGRFLWETYFSTADSIF
NTVLAMIRTRIQIPAAPIREEAAAEIPHRSGKAAGTDPNMADNGDLDLGPVETEPPYASPRYLRNFTLTVTDFYRSW
NCAPGPFHLFPHTPFDPVLPSEAKFLGSGTGFRPIGGGAGGSGKEFQAALGGNVPREQFTVVMLTYEREEVLMNSLE
RLNGLPYLNKVVVVWNSPKLPSEDLLWPDIGVPIMVVRTEKNSLNNRFLPWNEIETEAILSIDDDAHLRHDEIMFGF
RVWREARDRIVGFPGRYHAWDIPHQSWLYNSNYSCELSMVLTGAAFFHKYYAYLYSYVMPQAIRDMVDEYINCEDIA
MNFLVSHITRKPPIKVTSRWTFRCPGCPQALSHDDSHFHERHKCINFFVKVYGYMPLLYTQFRVDSVLFKTRLPHDK

TKCFKFI

<210> SEQ ID NO 1438
<211> Length : 957
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1438

>Kinesin heavy chain isoform 5C
MADPAECSIKVMCRFRPLNEAEILRGDKFIPKFKGDETVVIGQGKPYVFDRVLPPNTTQEQVYNACAKQIVKDVLEG
YNGTIFAYGQTSSGKTHTMEGKLHDPQLMGIIPRIAHDIFDHIYSMDENLEFHIKVSYFEIYLDKIRDLLDVSKTNL
AVHEDKNRVPYVKGCTERFVSSPEEVMDVIDEGKANRHVAVTNMNEHSSRSHSIFLINIKQENVETEKKLSGKLYLV
DLAGSEKVSKTGAEGAVLDEAKNINKSLSALGNVISALAEGTKTHVPYRDSKMTRILQDSLGGNCRTTIVICCSPSV
FNEAETKSTLMFGQRAKTIKNTVSVNLELTAEEWKKKYEKEKEKNKTLKNVIQHLEMELNRWRNGEAVPEDEQISAK
DQKNLEPCDNTPIIDNIAPVVAGISTEEKEKYDEEISSLYRQLDDKDDEINQQSQLAEKLKQQMLDQDELLASTRRD
YEKIQEELTRLQIENEAAKDEVKEVLQALEELAVNYDQKSQEVEDKTRANEQLTDELAQKTTTLTTTQRELSQLQEL
SNHQKKRATEILNLLLKDLGEIGGIIGTNDVKTLADVNGVIEEEFTMARLYISKMKSEVKSLVNRSKQLESAQMDSN
RKMNASERELAACQLLISQHEAKIKSLTDYMQNMEQKRRQLEESQDSLSEELAKLRAQEKMHEVSFQDKEKEHLTRL
QDAEEMKKALEQQMESHREAHQKQLSRLRDEIEEKQKIIDEIRDLNQKLQLEQEKLSSDYNKLKIEDQEREMKLEKL
LLLNDKREQAREDLKGLEETVSRELQTLHNLRKLFVQDLTTRVKKSVELDNDDGGGSAAQKQKISFLENNLEQLTKV
HKQLVRDNADLRCELPKLEKRLRATAERVKALESALKEAKENAMRDRKRYQQEVDRIKEAVRAKNMARRAHSAQIAK
PIRPGHYPASSPTAVHAIRGGGGSSSNSTHYQK

<210> SEQ ID NO 1439
<211> Length : 650
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1439

>Amyloid-like protein 1 precursor
MGPASPAARGLSRRPGQPPLPLLLPLLLLLLRAQPAIGSLAGGSPGAAEAPGSAQVAGLCGRLTLHRDLRTGRWEPD

PQRSRRCLRDPQRVLEYCRQMYPELQIARVEQATQAIPMERWCGGSRSGSCAHPHHQVVPFRCLPGEFVSEALLVPE
GCRFLHQERMDQCESSTRRHQEAQEACSSQGLILHGSGMLLPCGSDRFRGVEYVCCPPPGTPDPSGTAVGDPSTRSW
PPGSRVEGAEDEEEEESFPQPVDDYFVEPPQAEEEEETVPPPSSHTLAVVGKVTPTPRPTDGVDIYFGMPGEISEHE
GFLRAKMDLEERRMRQINEVMREWAMADNQSKNLPKADRQALNEHFQSILQTLEEQVSGERQRLVETHATRVIALIN
DQRRAALEGFLAALQADPPQAERVLLALRRYLRAEQKEQRHTLRHYQHVAAVDPEKAQQMRFQVHTHLQVIEERVNQ
SLGLLDQNPHLAQELRPQIQELLHSEHLGPSELEAPAPGGSSEDKGGLQPPDSKDDTPMTLPKGSTEQDAASPEKEK
MNPLEQYERKVNASVPRGFPFHSSEIQRDELAPAGTGVSREAVSGLLIMGAGGGSLIVLSMLLLRRKKPYGAISHGV
VEVDPMLTLEEQQLRELQRHGYENPTYRFLEERP

<210> SEQ ID NO 1440
<211> Length : 98
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1440

>Acylphosphatase, organ-common type isozyme
AEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHIDK
ANFNNEKVILKLDYSDFQIVK

<210> SEQ ID NO 1441
<211> Length : 99
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1441

>ACYO_HUMAN_V1
MAEGNTLISVDYEIFGKVQGVFFRKHTQAEGKKLGLVGWVQNTDRGTVQGQLQGPISKVRHMQEWLETRGSPKSHID
KANFNNEKVILKLDYSDFQIVK

<210> SEQ ID NO 1442
<211> Length : 246
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1442

>Sorting nexin 26
MLSLSLCSHLWGPLILSALQARSTDSLDGPGEGSVQPLPTAGGPSVKGKPGKRLSAPRGPFPRLADCAHFHYENVDF
GHIQLLLSPDREGPSLSGENELVFGVQVTCQGRSWPVLRSYDDFRSLDAHLHRCIFDRRFSCLPELPPPPEGARAAQ
MLVPLLLQYLETLSGLVDSNLNCGPVLTWMEVGLGRGLGDSEWVRGCVCHHAQHREILDGNRVASAVEDEGAEVDGE
AFRWGSLWVGESWDM

<210> SEQ ID NO 1443
<211> Length : 862
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1443

>Q9NT23
HDVIQQLPPPHYRTLEYLLRHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFL
LTHVDVLFSDTFTSAGLDPAGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGER
GEKQRKPGGSSWKTFFALGRGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLR
RPHSSSDAFPVGPAPAGSCESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLE
GLRGLDFDPLTFRCSSPTPGDPAPPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLEL
LGAGGAPASATPTPALSPGRSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGG
APPPPPKNPARLMALALAERAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVP
GPPPYLPRQQSDGSLLRSQRPMGTSRRGLRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECL
PPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYEIGASEGSPYSGLTRSWSPFRSMPPDRLNAS
YGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRS
RSDPGPPVPRLPQKQRAPWGPRTPHRVPGPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPT
PSWSLHSEGQTRSYC

<210> SEQ ID NO 1444

<211> Length : 295
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1444

>Q96CP3
LRGPAQVSAQLRAGGGGRDAPEAAAQSPCSVPSQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVW
RSSLGPPAPLDRGENLYYEIGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFP
PDHLGYSAPQHPARRPTPPEPLYVNLALGPRGPSPASSSSSSPPAHPRSRSDPGPPVPRLPQKQRAPWGPRTPHRVP
GPWGPPEPLLLYRAAPPAYGRGGELHRGSLYRNGGQRGEGAGPPPPYPTPSWSLHSEGQTRSYC

<210> SEQ ID NO 1445
<211> Length : 1007
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1445

>BAC86902
MLVPLLLQYLETLSGLVDSNLNCGPVLTWMELDNHGRRLLLSEEASLNIPAVAAAHVIKRYTAQAPDELSFEVGDIV
SVIDMPPTEDRSWWRGKRGFQVGFFPSECVELFTERPGPGLKADADGPPCGIPAPQGISSLTSAVPRPRGKLAGLLR
TFMRSRPSRQRLRQRGILRQRVFGCDLGEHLSNSGQDVPQVLRCCSEFIEAHGVVDGIYRLSGVSSNIQRLRHEFDS
ERIPELSGPAFLQDIHSVSSLCKLYFRELPNPLLTYQLYGKFSEAMSVPGEEERLVRVHDVIQQLPPPHYRTLEYLL
RHLARMARHSANTSMHARNLAIVWAPNLLRSMELESVGMGGAAAFREVRVQSVVVEFLLTHVDVLFSDTFTSAGLDP
AGRCLLPRPKSLAGSCPSTRLLTLEEAQARTQGRLGTPTEPTTPKAPASPAERRKGERGEKQRKPGGSSWKTFFALG
RGPSVPRKKPLPWLGGTRAPPQPSGSRPDTVTLRSAKSEESLSSQASGAGLQRLHRLRRPHSSSDAFPVGPAPAGSC
ESLSSSSSSESSSSESSSSSSESSAAGLGALSGSPSHRTSAWLDDGDELDFSPPRCLEGLRGLDFDPLTFRCSSPTP
GDPAPPASPAPPAPASAFPPRVTPQAISPRGPTSPASPAALDISEPLAVSVPPAVLELLGAGGAPASATPTPALSPG
RSLRPHLIPLLLRGAEAPLTDACQQEMCSKLRGAQGPLGPDMESPLPPPPLSLLRPGGAPPPPPKNPARLMALALAE
RAQQVAEQQSQQECGGTPPASQSPFHRSLSLEVGGEPLGTSGSGPPPNSLAHPGAWVPGPPPYLPRQQSDGSLLRSQ
RPMGTSRRGLRGPAQVPTPGFFSPAPRECLPPFLGVPKPGLYPLGPPSFQPSSPAPVWRSSLGPPAPLDRGENLYYE
IGASEGSPYSGPTRSWSPFRSMPPDRLNASYGMLGQSPPLHRSPDFLLSYPPAPSCFPPDHLGYSPPSTLLGALHRL
SPSTST

<210> SEQ ID NO 1446
<211> Length : 1806
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1446

>Collagen alpha 1
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEANIVDDFQE
YNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYE
YKEYEDKPTSPPNEEFGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPT
GPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGR
PGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDRGFDGLPGLPGDKGHRG
ERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGPRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQ
GNPGPQGLPGPQGPIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPPGPQGPIGYPGPRGVKGADGV
RGLKGSKGEKGEDGFPGFKGDMGLKGDRGEVGQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLPGYPG
RQGPKGSTGFPGFPGANGEKGARGVAGKPGPRGQRGPTGPRGSRGARGPTGKPGPKGTSGGDGPPGPPGERGPQGPQ
GPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGYPGPPGPPGEQGLPGA
AGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQGPPGPVGSPGERGSAGTAGPIGLRGRPG
PQGPPGPAGEKGAPGEKGPQGPAGRDGVQGPVGLPGPAGPAGSPGEDGDKGEIGEPGQKGSKGGKGENGPPGPPGLQ
GPVGAPGIAGGDGEPGPRGQQGMFGQKGDEGARGFPGPPGPIGLQGLPGPPGEKGENGDVGPMGPPGPPGPRGPQGP
NGADGPQGPPGSVGSVGGVGEKGEPGEAGNPGPPGEAGVGGPKGERGEKGEAGPPGAAGPPGAKGPPGDDGPKGNPG
PVGFPGDPGPPGELGPAGQDGVGGDKGEDGDPGQPGPPGPSGEAGPPGPPGKRGPPGAAGAEGRQGEKGAKGEAGAE
GPPGKTGPVGPQGPAGKPGPEGLRGIPGPVGEQGLPGAAGQDGPPGPMGPPGLPGLKGDPGSKGEKGHPGLIGLIGP
PGEQGEKGDRGLPGTQGSPGAKGDGGIPGPAGPLGPPGPPGLPGPQGPKGNKGSTGPAGQKGDSGLPGPPGPPGPPG
EVIQPLPILSSKKTRRHTEGMQADADDNILDYSDGMEEIFGSLNSLKQDIEHMKFPMGTQTNPARTCKDLQLSHPDF
PDGEYWIDPNQGCSGDSFKVYCNFTSGGETCIYPDKKSEGVRISSWPKEKPGSWFSEFKRGKLLSYLDVEGNSINMV
QMTFLKLLTASARQNFTYHCHQSAAWYDVSSGSYDKALRFLGSNDEEMSYDNNPFIKTLYDGCTSRKGYEKTVIEIN
TPKIDQVPIVDVMISDFGDQNQKFGFEVGPVCFLG

<210> SEQ ID NO 1447
<211> Length : 1806

<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1447

>CA1B_HUMAN_V5
MEPWSSRWKTKRWLWDFTVTTLALTFLFQAREVRGAAPVDVLKALDFHNSPEGISKTTGFCTNRKNSKGSDTAYRVS
KQAQLSAPTKQLFPGGTFPEDFSILFTVKPKKGIQSFLLSIYNEHGIQQIGVEVGRSPVFLFEDHTGKPAPEDYPLF
RTVNIADGKWHRVAISVEKKTVTMIVDCKKKTTKPLDRSERAIVDTNGITVFGTRILDEEVFEGDIQQFLITGDPKA
AYDYCEHYSPDCDSSAPKAAQAQEPQIDEYAPEDIIEYDYEYGEAEYKEAESVTEGPTVTEETIAQTEANIVDDFQE
YNYGTMESYQTEAPRHVSGTNEPNPVEEIFTEEYLTGEDYDSQRKNSEDTLYENKEIDGRDSDLLVDGDLGEYDFYE
YKEYEDKPTSPPNEEFGPGVPAETDITETSINGHGAYGEKGQKGEPAVVEPGMLVEGPPGPAGPAGIMGPPGLQGPT
GPPGDPGDRGPPGRPGLPGADGLPGPPGTMLMLPFRYGGDGSKGPTISAQEAQAQAILQQARIALRGPPGPMGLTGR
PGPVGGPGSSGAKGESGDPGPQGPRGVQGPPGPTGKPGKRGRPGADGGRGMPGEPGAKGDRGFDGLPGLPGDKGHRG
ERGPQGPPGPPGDDGMRGEDGEIGPRGLPGEAGPRGLLGPRGTPGAPGQPGMAGVDGPPGPKGNMGPQGEPGPPGQQ
GNPGPQGLPGPQGPIGPPGEKGPQGKPGLAGLPGADGPPGHPGKEGQSGEKGALGPPGPQGPIGYPGPRGVKGADGV
RGLKGSKGEKGEDGFPGFKGDMGLKGDRGEVGQIGPRGEDGPEGPKGRAGPTGDPGPSGQAGEKGKLGVPGLPGYPG
RQGPKGSTGFPGFPGANGEKGARGVAGKPGPRGQRGPTGPRGSRGARGPTGKPGPKGTSGGDGPPGPPGERGPQGPQ
GPVGFPGPKGPPGPPGKDGLPGHPGQRGETGFQGKTGPPGPGGVVGPQGPTGETGPIGERGHPGPPGPPGEQGLPGA
AGKEGAKGDPGPQGISGKDGPAGLRGFPGERGLPGAQGAPGLKGGEGPQGPPGPVGSPGERGSAGTAGPIGLRGRPG
PQGPPGPAGEKGAPGEKGPQGPAGRDGVQGPVGLPGPAGPAGSPGEDGDKGEIGEPGQKGSKGGKGENGPPGPPGLQ
GPVGAPGIAGGDGEPGPRGQQGMFGQKGDEGARGFPGPPGPIGLQGLPGPPGEKGENGDVGPMGPPGPPGPRGPQGP
NGADGPQGPPGSVGSVGGVGEKGEPGEAGNPGPPGEAGVGGPKGERGEKGEAGPPGAAGPPGAKGPPGDDGPKGNPG
PVGFPGDPGPPGELGPAGQDGVGGDKGEDGDPGQPGPPGPSGEAGPPGPPGKRGPPGAAGAEGRQGEKGAKGEAGAE
GPPGKTGPVGPQGPAGKPGPEGLRGIPGPVGEQGLPGAAGQDGPPGPMGPPGLPGLKGDPGSKGEKGHPGLIGLIGP
PGEQGEKGDRGLPGTQGSPGAKGDGGIPGPAGPLGPPGPPGLPGPQGPKGNKGSTGPAGQKGDSGLPGPPGPPGPPG
EVIQPLPILSSKKTRRHTEGMQADADDNILDYSDGMEEIFGSLNSLKQDIEHMKFPMGTQTNPARTCKDLQLSHPDF
PDGEYWIDPNQGCSGDSFKVYCNFTSGGETCIYPDKKSEGVRISSWPKEKPGSWFSEFKRGKLLSYLDVEGNSINMV
QMTFLKLLTASARQNFTYHCHQSAAWYDVSSGSYDKALRFLGSNDEEMSYDNNPFIKTLYDGCTSRKGYEKTVIEIN
TPKIDQVPIVDVMISDFGDQNQKFGFEVGPVCFLG

<210> SEQ ID NO 1448
<211> Length : 153
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1448

>Myoglobin
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILK
KKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1449
<211> Length : 154
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1449

>MYG_HUMAN_V1
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1450
<211> Length : 99
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1450

>Q8WVH6
MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAM
NKALELFRKDMASNYKELGFQG

<210> SEQ ID NO 1451
<211> Length : 702
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1451

>Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
MESPSAPPHRWCIPWQRLLLTASLLTFWNPPTTAKLTIESTPFNVAEGKEVLLLVHNLPQHLFGYSWYKGERVDGNR
QIIGYVIGTQQATPGPAYSGREIIYPNASLLIQNIIQNDTGFYTLHVIKSDLVNEEATGQFRVYPELPKPSISSNNS
KPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVTRNDTASYKCETQNPVSARRSDSVIL
NVLYGPDAPTISPLNTSYRSGENLNLSCHAASNPPAQYSWFVNGTFQQSTQELFIPNITVNNSGSYTCQAHNSDTGL
NRTTVTTITVYAEPPKPFITSNNSNPVEDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTR
NDVGPYECGIQNELSVDHSDPVILNVLYGPDDPTISPSYTYYRPGVNLSLSCHAASNPPAQYSWLIDGNIQQHTQEL
FISNITEKNSGLYTCQANNSASGHSRTTVKTITVSAELPKPSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQ
SLPVSPRLQLSNGNRTLTLFNVTRNDARAYVCGIQNSVSANRSDPVTLDVLYGPDTPIISPPDSSYLSGANLNLSCH
SASNPSPQYSWRINGIPQQHTQVLFIAKITPNNNGTYACFVSNLATGRNNSIVKSITVSASGTSPGLSAGATVGIMI
GVLVGVALI

<210> SEQ ID NO 1452
<211> Length : 495
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1452

>Alanine aminotransferase
ASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQRP
ITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNVF
LSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELARALGQARDHC
RPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPYAGQQELASF
HSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCPPVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAEL
AAKAKLTEQVFNEAPGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGICVVPGSGFGQREGT
YHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS

<210> SEQ ID NO 1453
<211> Length : 496
<212> Type : PRT

<213> Organism : Homo sapiens

<400> sequence : 1453

>ALAT_HUMAN_V1
MASSTGDRSQAVRHGLRAKVLTLDGMNPRVRRVEYAVRGPIVQRALELEQELRQGVKKPFTEVIRANIGDAQAMGQR
PITFLRQVLALCVNPDLLSSPNFPDDAKKRAERILQACGGHSLGAYSVSSGIQLIREDVARYIERRDGGIPADPNNV
FLSTGASDAIVTVLKLLVAGEGHTRTGVLIPIPQYPLYSATLAELGAVQVDYYLDEERAWALDVAELHRALGQARDH
CRPRALCVINPGNPTGQVQTRECIEAVIRFAFEERLFLLADEVYQDNVYAAGSQFHSFKKVLMEMGPPYAGQQELAS
FHSTSKGYMGECGFRGGYVEVVNMDAAVQQQMLKLMSVRLCPPVPGQALLDLVVSPPAPTDPSFAQFQAEKQAVLAE
LAAKAKLTEQVFNEAPGISCNPVQGAMYSFPRVQLPPRAVERAQELGLAPDMFFCLRLLEETGICVVPGSGFGQREG
TYHFRMTILPPLEKLRLLLEKLSRFHAKFTLEYS

<210> SEQ ID NO 1454
<211> Length : 132
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1454

>Antileukoproteinase 1 precursor
MKSSGLFPFLVLLALGTLAPWAVEGSGKSFKAGVCPPKKSAQCLRYKKPECQSDWQCPGKKRCCPDTCGIKCLDPVD
TPNPTRRKPGKCPVTYGQCLMLNPPNFCEMDGQCKRDLKCCMGMCGKSCVSPVKA

<210> SEQ ID NO 1455
<211> Length : 488
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1455

>Stromelysin-3 precursor
MAPAAWLRSAAARALLPPMLLLLLQPPPLLARALPPDVHHLHAERRGPQPWHAALPSSPAPAPATQEAPRPASSLRP
PRCGVPDPSDGLSARNRQKRFVLSGGRWEKTDLTYRILRFPWQLVQEQVRQTMAEALKVWSDVTPLTFTEVHEGRAD
IMIDFARYWDGDDLPFDGPGGILAHAFFPKTHREGDVHFDYDETWTIGDDQGTDLLQVAAHEFGHVLGLQHTTAAKA

LMSAFYTFRYPLSLSPDDCRGVQHLYGQPWPTVTSRTPALGPQAGIDTNEIAPLEPDAPPDACEASFDAVSTIRGEL
FFFKAGFVWRLRGGQLQPGYPALASRHWQGLPSPVDAAFEDAQGHIWFFQGAQYWVYDGEKPVLGPAPLTELGLVRF
PVHAALVWGPEKNKIYFFRGRDYWRFHPSTRRVDSPVPRRATDWRGVPSEIDAAFQDADGYAYFLRGRLYWKFDPVK
VKALEGFPRLVGPDFFGCAEPANTFL

<210> SEQ ID NO 1456
<211> Length : 80
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1456

>Trefoil factor 3 precursor
MAARALCMLGLVLALLSSSSAEEYVGLSANQCAVPAKDRVDCGYPHVTPKECNNRGCCFDSRIPGVPWCFKPLQEAE
CTF

<210> SEQ ID NO 1457
<211> Length : 95
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1457

>S-100P protein
MTELETAMGMIIDVFSRYSGSEGSTQTLTKGELKVLMEKELPGFLQSGKDKDAVDKLLKDLDANGDAQVDFSEFIVF
VAAITSACHKYFEKAGLK

<210> SEQ ID NO 1458
<211> Length : 302
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1458

>Stanniocalcin 2 precursor
MCAERLGQFMTLALVLATFDPARGTDATNPPEGPQDRSSQQKGRLSLQNTAEIQHCLVNAGDVGCGVFECFENNSCE
IRGLHGICMTFLHNAGKFDAQGKSFIKDALKCKAHALRHRFGCISRKCPAIREMVSQLQRECYLKHDLCAAAQENTR
VIVEMIHFKDLLLHEPYVDLVNLLLTCGEEVKEAITHSVQVQCEQNWGSLCSILSFCTSAIQKPPTAPPERQPQVDR
TKLSRAHHGEAGHHLPEPSSRETGRGAKGERGSKSHPNAHARGRVGGLGAQGPSGSSEWEDEQSEYSDIRR

<210> SEQ ID NO 1459
<211> Length : 578
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1459

>Putative alpha-mannosidase C20orf31 precursor
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT
LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVS
MPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATG
DPTLLELGRDAVESIEKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLFDPTNFIHNNGSTFDAVITPYG
ECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVEDLMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPEN
HDQARERKPAKQKVPLLSCPSQPFTSKLALLGQVFLDSS

<210> SEQ ID NO 1460
<211> Length : 578
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1460

>AAH16184
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYRERVKAMFYHAYDSYLENAFPFDELRPLTCDGHDTWGSFSLT

LIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNASVFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMA
EEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTCTAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSD
IGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGAILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVS
MPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQFGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATG
DPTLLELGRDAVESIEKISKVECGFATIKDLRDHKLDNRMESFFLAETVKYLYLLLFDPTNFIHNNGSTFDTVITPYG
ECILGAGGYIFNTEAHPIDPAALHCCQRLKEEQWEVEDLMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPEN
HDQARERKPAKQKVPLLSCPSQPFTSKLALLGQVFLDSS

<210> SEQ ID NO 1461
<211> Length : 541
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1461

>AAQ88943
MPFRLLIPLGLLCALLPQHHGAPGPDGSAPDPAHYSFSLTLIDALDTLLILGNVSEFQRVVEVLQDSVDFDIDVNAS
VFETNIRVVGGLLSAHLLSKKAGVEVEAGWPCSGPLLRMAEEAARKLLPAFQTPTGMPYGTVNLLHGVNPGETPVTC
TAGIGTFIVEFATLSSLTGDPVFEDVARVALMRLWESRSDIGLVGNHIDVLTGKWVAQDAGIGAGVDSYFEYLVKGA
ILLQDKKLMAMFLEYNKAIRNYTRFDDWYLWVQMYKGTVSMPVFQSLEAYWPGLQSLIGDIDNAMRTFLNYYTVWKQ
FGGLPEFYNIPQGYTVEKREGYPLRPELIESAMYLYRATGDPTLLELGRDAVESIEKISKVECGFATIKDLRDHKLD
NRMESFFLAETVKYLYLLLFDPTNFIHNNGSTFDAVITPYGECILGAGGYIFNTEAHPIDLAALHCCQRLKEEQWEVE
DLMREFYSLKRSRSKFQKNTVSSGPWEPPARPGTLFSPENHDQARERKPAKQKVPLLSCPSQPFTSKLALLGQVFLD
SS

<210> SEQ ID NO 1462
<211> Length : 314
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1462

>Osteopontin precursor
MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQKQNLLAPQNAVSSEETNDFKQETLPSK

SNESHDHMDDMDDEDDDDHVDSQDSIDSNDSDDVDDTDDSHQSDESHHSDESDELVTDFPTDLPATEVFTPVVPTVD
TYDGRGDSVVYGLRSKSKKFRRPDIQYPDATDEDITSHMESEELNGAYKAIPVAQDLNAPSDWDSRGKDSYETSQLD
DQSAETHSHKQSRLYKRKANDESNEHSDVIDSQELSKVSREFHSHEFHSHEDMLVVDPKSKEEDKHLKFRISHELDS
ASSEVN

<210> SEQ ID NO 1463
<211> Length : 357
<212> Type : PRT
<213> Organism : Homo sapiens

<400> sequence : 1463

>NOV protein homolog precursor
MQSVQSTSFCLRKQCLCLTFLLLHLLGQVAATQRCPPQCPGRCPATPPTCAPGVRAVLDGCSCCLVCARQRGESCSD
LEPCDESSGLYCDRSADPSNQTGICTAVEGDNCVFDGVIYRSGEKFQPSCKFQCTCRDGQIGCVPRCQLDVLLPEPN
CPAPRKVEVPGECCEKWICGPDEEDSLGGLTLAAYRPEATLGVEVSDSSVNCIEQTTEWTACSKSCGMGFSTRVTNR
NRQCEMLKQTRLCMVRPCEQEPEQPTDKKGKKCLRTKKSLKAIHLQFKNCTSLHTYKPRFCGVCSDGRCCTPHNTKT
IQAEFQCSPGQIVKKPVMVIGTCTCHTNCPKNNEAFLQELELKTTRGKM

<210> SEQ ID NO 1464
<211> Length : 516
<212> Type : DNA
<213> Organism : Homo sapiens

<400> sequence : 1464

>HSU33147_PEA_1_T1
GTGCTCACCTCCACAGCGGCTTCCTTGATCCTTGCCACCCGCGACTGAACACCGACAGCAGCAGCCTCACCATGAAG
TTGCTGATGGTCCTCATGCTGGCGGCCCTCTCCCAGCACTGCTACGCAGGCTCTGGCTGCCCCTTATTGGAGAATGT
GATTTCCAAGACAATCAATCCACAAGTGTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGACGACAATG
CCACTACAAATGCCATAGATGAATTGAAGGAATGTTTTCTTAACCAAACGGATGAAACTCTGAGCAATGTTGAGCAA

TTAATATATGACAGCAGTCTTTGTGATTTATTTTAACTTTCTGCAAGACCTTTGGCTCACAGAACTGCAGGGTATGG
TGAGAAACCAACTACGGATTGCTGCAAACCACACCTTCTCTTTCTTATGTCTTTTTACTACAAACTACAAGACAATT
GTTGAAACCTGCTATACATGTTTATTTTAATAAATTGATGGCAAAAAAAAAAAT

<210> SEQ ID NO 1465

<211> Length : 907

<212> Type : DNA

<213> Organism : Homo sapiens

<400> sequence : 1465

>HSU33147_PEA_1_T2
GTGCTCACCTCCACAGCGGCTTCCTTGATCCTTGCCACCCGCGACTGAACACCGACAGCAGCAGCCTCACCATGAAG
TTGCTGATGGTCCTCATGCTGGCGGCCCTCTCCCAGCACTGCTACGCAGGCTCTGGCTGCCCCTTATTGGAGAATGT
GATTTCCAAGACAATCAATCCACAAGTGTCTAAGACTGAATACAAAGAACTTCTTCAAGAGTTCATAGACGACAATG
CCACTACAAATGCCATAGATGAATTGAAGGAATGTTTTCTTAACCAAACGGATGAAACTCTGAGCAATGTTGAGGTG
TTTATGGTAATTTCATTTTCTTCCTATAAGCTTTTTAAATCCCCTGACCAGGGACAAGTGGGCTCTTCATTTCTCAC
TGACAATGCCAAAGCCACTAGTGAACAAGCCTTTTCTTACATTGGTTAATTTAGTTGAATGGTTAGTCTAATGACTT
TGCCATCAAGAAAAACATCCAGTGTCCCTGTGTTGTCACTCTACCCAGAGAATCCTCAGTGGATGATAAATGAATAG
GGCAAGAGAGGAAAAGGAAAGGTCGGTAGAAGTCTTACCTATCCCCAGAGCTCTCTAATTCATGCTCACAAACACAG
ACACAATCACACAAACACAGAAACACACATACACACATCCAGACACATGCAAACACACAGACACAGTCACAATCACA
CAAACACACACACATTCAGACATACACAAACATAGACAGACAGGCAAAGACACAGACACAGACACAGACACAATCAC
ACCAGCACACAATCATCCAGACACAAACACAAACACACAGACAGAACCACACAACCACAGAAACACAGAGACACACA
CAAACACACTCAGACACACACATACAAACATATGTTCACTCTCTACAGAAAAAACAATTT

SEQ ID NO:1466
>HPRT1-F:
TGACACTGGCAAAACAATGCA

SEQ ID NO:1467
>HPRT1-R:
GGTCCTTTTCACCAGCAAGCT

SEQ ID NO:1468
>amplicon

TGACACTGGCAAAACAATGCAGACTTTGCTTTCCTTGGTCAGGCA
GTATAATCCAAAGATGGTCAAGGTCGCAAGCTTGCTGGTGAAAAGGACC

SEQ ID NO:1469
>PBGD-F2:
TGAGAGTGATTCGCGTGGG

SEQ ID NO:1470
>PBGD-R2:
CCAGGGTACGAGGCTTTCAAT

SEQ ID NO:1471
>amplicon
TGAGAGTGATTCGCGTGGGTACCCGCAAGAGCCAGCTTGCTCGCATACAGACGGACAGTGTGGTGGCAACATTGAAA
GCCTCGTACCCTGG

SEQ ID NO:1472
>Ubiquitin-F:
ATTTGGGTCGCGGTTCTTG

SEQ ID NO:1473
>Ubiquitin-R:
TGCCTTGACATTCTCGATGGT

SEQ ID NO:1474
>Amplicon:
ATTTGGGTCGCGGTTCTTGTTTGTGGATCGCTGTGATCGTCACTTGACAATGCAGATCTTCGTGAAGACTCTGACTG
GTAAGACCATCACCCTCGAGG TTGAGCCCAGTGACACCATCGAGAATGTCAAGGCA

SEQ ID NO:1475
>SDHA-F:
TGGGAACAAGAGGGCATCTG

SEQ ID NO:1476
>SDHA-R:
CCACCACTGCATCAAATTCATG

SEQ ID NO:1477
>amplicon
TGGGAACAAGAGGGCATCTGCTAAAGTTTCAGATTCCATTTCT
GCTCAGTATCCAGTAGTGGATCATGAATTTGATGCAGTGGTGG

SEQ ID NO:1478
>Forward primer:
AGACTCCAACCCACAGCCC

SEQ ID NO:1479
>Reverse primer:
CAGCTCAGCCAACCTTGCA

SEQ ID NO:1480
>Amplicon:
AGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGGGGGCCACCT
GGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTG

SEQ ID NO:1481
>T86235 # transcript_8 #len 1491 (Includes node 44 - TAA seg 35)
CTGTCTCCACTAAAATTTTAAAAATTAGAAAGTGCTCTCTGGAAAAGCTGCCTAACTCTCACTGCTTCTCTGCTGCC
CCTCCTAATGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGGGTGGGACTCAGGCTGGTCT
TTCTCCTGCCCCAGCCTGGCTTGCTTGTGTGCCTTGTTCCTTGGTGACAGGAGCAGGTTGCCGTCCGGTTGTTTGAC
CAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAG
AACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGC
TGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAG
ATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCT
GCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCT
TCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCA
GAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGT
ACCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTGAGCCTGAGATACCGGAGTCCT
CTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGG
CAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCT
AGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGC
TTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGCCTCAGCAATCTGGCC
CCTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGA

CGATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACAT
TACTCGAATGGCAGGATGCCCTGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAAC
CACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCAC
TTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1482
>T86235 # transcript_9 #len 1644 (Includes node 44 - TAA seg 35 ; node 32 - TAA seg 37)
CTGTCTCCACTAAAATTTTAAAAATTAGAAAGTGCTCTCTGGAAAAGCTGCCTAACTCTCACTGCTTCTCTGCTGCC
CCTCCTAATGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGGGTGGGACTCAGGCTGGTCT
TTCTCCTGCCCCAGCCTGGCTTGCTTGTGTGCCTTGTTCCTTGGTGACAGGAGCAGGTTGCCGTCCGGTTGTTTGAC
CAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAG
AACCCCCAGCCTCCAGGAGGTGAAGATTCAAGTGAGTCTGTGTGGCCAACAGCTTTGATGTCTATTGAACAGTGACT
GGGCTGAGGAAGAGGGAAAAGAGATGGGGGATCAGGAATAGGACAGTGTGGGTAGACTACTGAACGCACATCTTGAT
GTCACACTGGGGTGCTCTCTCCCACCACAGCGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCT
GGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGA
TTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTG
CCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTT
CTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAG
AACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTA
CCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTGAGCCTGAGATACCGGAGTCCTC
TCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGC
AGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTA
GAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGCT
TTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGCCTCAGCAATCTGGCCC
CTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGAC
GATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATT
ACTCGAATGGCAGGATGCCCTGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACC
ACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACT
TTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1483
>T86235 # transcript_10 #len 1404 (Includes node 44 - TAA seg 35)
CTGTCTCCACTAAAATTTTAAAAATTAGAAAGTGCTCTCTGGAAAAGCTGCCTAACTCTCACTGCTTCTCTGCTGCC
CCTCCTAATGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGGGTGGGACTCAGGCTGGTCT
TTCTCCTGCCCCAGCCTGGCTTGCTTGTGTGCCTTGTTCCTTGGTGACAGGAGCAGGTTGCCGTCCGGTTGTTTGAC
CAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAG
AACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGC

TGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAG
ATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCT
GCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCT
TCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCA
GAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGA
ACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGGAC
CCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCAGT
CTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCCAG
CCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGCCTCAGCAATCTGGCCCCTCGAACCC
TAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGT
GCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATG
GCAGGATGCCCTGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCTGCC
CTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTCCTC
AATAAAGTTTCTAAAGTA

SEQ ID NO:1484

>T86235 # transcript_22 #len 2797 (Includes node 37 - TAA seg 42)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGGTGGGGATCAGTCAGCCTCGGAACCCCTTGGAAGAGCTCAGGCCTAGCCCTAGGGGTCAAAAT
GTGGGGCCTGGGCCCCCTGCCCAGACAGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCAC
CATCCTGTCAGGTGAGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAG
GAAGTCAGGGAGGCACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCAC
CGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCA
GAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAA
GGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCT
TTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCA
GCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTG
CCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCC
CCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCAT
GCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCC
TGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCA
CCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTAT
CCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTC
TGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACT
TCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCC

CTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGG
AACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAG
ATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTC
CTACTGTAGGATTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAG
CAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGG
GCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGC
AACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGAC
CCCCAGCAGGCCAGGCAGGTAAGGAGTTGGCTGGGAAGGAGTGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTG
CACACCTGCCCTGCCCCTACCCCAGGCAATCTCATGCTTCCACACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCA
GGAAGAGGGGAGGGGCTGCACTTCCAGCCCTGTGCTCCTAATTGGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACA
GTACATGGTGGAAGTATAGGACCCCAGACCTCCCTCTAAATTTTCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCT
CGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGA
TGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTAC
TCGAATGGCAGGATGCCCTGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCAC
TCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTT
TGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1485
>T86235 # transcript_23 #len 2962 (Includes node 39 - TAA seg 44; node 37 - TAA seg 42)
CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGGTGGGGATCAGTCAGCCTCGGAACCCCTTGGAAGAGCTCAGGCCTAGCCCTAGGGGTCAAAAT
GTGGGGCCTGGGCCCCCTGCCCAGACAGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCAC
CATCCTGTCAGGTGAGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAG
GAAGTCAGGGAGGCACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCAC
CGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCA
GAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAA
GGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCT
TTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCA
GCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTG
CCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCC
CCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCAT
GCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCC
TGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCA
CCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTAT

CCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTC
TGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACT
TCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCC
CTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGG
AACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAG
ATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTC
CTACTGTAGGATTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAG
CAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGG
GCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGC
AACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGAC
CCCCAGCAGGCCAGGCAGGTAAGGAGTTGGCTGGGAAGGAGTGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTG
CACACCTGCCCTGCCCCTACCCCAGGCAATCTCATGCTTCCACACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCA
GGAAGAGGGGAGGGGCTGCACTTCCAGCCCTGTGCTCCTAATTGGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACA
GTACATGGTGGAAGTATAGGACCCCAGACCTCCCTCTAAATTTTCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCT
CGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGA
TGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTAC
TCGAATGGCAGGATGCCCTGGTGAGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGG
GGAACAGGGACAGGGGGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGC
TGAGCTGTGACAAGGTCTTCTCTGTCTCCAGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGAT
GAGACAACCACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAG
CCGCCCACTTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1486

>T86235 # transcript_24 #len 2605 (Includes node 39 - TAA seg 44)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGGTGGGGATCAGTCAGCCTCGGAACCCCTTGGAAGAGCTCAGGCCTAGCCCTAGGGGTCAAAAT
GTGGGGCCTGGGCCCCCTGCCCAGACAGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCAC
CATCCTGTCAGGTGAGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAG
GAAGTCAGGGAGGCACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCAC
CGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCA
GAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAA
GGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCT
TTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCA
GCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTG
CCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCC

CCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCAT
GCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCC
TGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCA
CCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTAT
CCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTC
TGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACT
TCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCC
CTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGG
AACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAG
ATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGG
GCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGC
CCTGCACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGC
CTGATCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGG
CCAGGCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCC
ACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGG
GTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGATGCCCTGGTGAGACTCCAACCCACAGCCCAGCTGTG
GCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGGGGGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAG
GAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTGTGACAAGGTCTTCTCTGTCTCCAGTGTTTCATTCCAGTTGGT
TCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTT
ATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1487

>T86235 # transcript_25 #len 2254 (Includes node 39 - TAA seg 44)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGGGTGTTCGGGC
CTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGG
GACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCAC
CCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTC
AGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAG
ATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGG
GACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCC
ATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTC
TCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGT
GTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAG
TTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCA
GCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGG
GGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAG

AACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGC
CCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGG
AGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAG
GCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGAACAGCTTG
AGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGG
CCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACC
CTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTA
TCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCTG
AGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTTA
CACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGATG
CCCTGGTGAGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGGG
GGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTGTGACAAGG
TCTTCTCTGTCTCCAGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCT
GCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTC
CTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1488

>T86235 # transcript_26 #len 2611 (Includes node 39 - TAA seg 44; node 37 - TAA seg 42)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGGGTGTTCGGGC
CTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGG
GACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCAC
CCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTC
AGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAG
ATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGG
GACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCC
ATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTC
TCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGT
GTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAG
TTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCA
GCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGG
GGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAG
AACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGC
CCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGG
AGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAG
GCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGC

CCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTGAGCCTGAGATACCGGAGTCCTCTCGCCAG
GAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGG
ACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCA
GTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCC
AGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGTAAGGAGTTGGCTGGGAAGGAGTG
TGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGCCCCTACCCCAGGCAATCTCATGCTTCCAC
ACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCAGGAAGAGGGGAGGGGCTGCACTTCCAGCCCTGTGCTCCTAATT
GGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACAGTACATGGTGGAAGTATAGGACCCCAGACCTCCCTCTAAATTT
TCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCG
CCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCC
ACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGATGCCCTGGTGAGACTCCAACCCACAGCCCA
GCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGGGGGCCACCTGGGCTTCTTCACAGAGAGGT
CAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTGTGACAAGGTCTTCTCTGTCTCCAGTGTTTCATTCCA
GTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCC
TTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1489

>T86235 # transcript_27 #len 2446 (Includes node 37 - TAA seg 42)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGGGTGTTCGGGC
CTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGCCAGGGCTTCCTGCTTCTCTAGGCTGGAGG
GACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTCTGATTTCACCTTCAGGACCTTCCTTTCAC
CCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGCAGCAGCCGGACTTCAGTGAGCCAGGCCTC
AGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAAAGGAGAACGCGAGGTTGTCACTCACTCAG
ATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAAGAGAAAACCGAGAAATGTCACATACCAGG
GACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAGCCCTTGCCTGGCCATGTGGTGCCATGTCC
ATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAACTCTGACCTCATATTCAGTGTTGCGGCGTC
TCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCAGAGTTCAGCAGGCCCAGTGGCTGCGTGGT
GTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAGGTTGCCGTCCGGTTGTTTGACCAGGAGAG
TTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCCTTCTGGACCCCACTCTAACAGAACCCCCA
GCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGTTGAGACAGGAAGTAGAGGGGCTGGTAGGG
GGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAACTTCAGCCCCTGCTGACTGAGATTTCTAG
AACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGGACTGTTAAAACACTCAGGGCTGCCAAAGC
CCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAGAGCCTGGGCCCCCAGAGGCCTTCTGTAGG
AGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTACCAGAGCCCTACCCTCCAGCAGAACCCAG
GCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGC
CCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTGAGCCTGAGATACCGGAGTCCTCTCGCCAG

GAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCAGCACCCAGGGGCAGTCTGG
ACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAACATAGAAGTCTAGAGTCCA
GTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTCCCAACACCCGCTTTGTGCC
AGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGTAAGGAGTTGGCTGGGAAGGAGTG
TGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGCCCCTACCCCAGGCAATCTCATGCTTCCAC
ACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCAGGAAGAGGGGAGGGGCTGCACTTCCAGCCCTGTGCTCCTAATT
GGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACAGTACATGGTGGAAGTATAGGACCCCAGACCTCCCTCTAAATTT
TCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCG
CCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCC
ACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGATGCCCTGTGTTTCATTCCAGTTGGTTCTGC
TGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGT
CGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1490
>T86235 # transcript_29 #len 844 (Includes node 11 - TAA seg 14; node 6 - TAA seg 9)
CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGGTAAGAGGGGC
CTAATGGGGGAAGACAGTAGTCACACCAGTAATGCACCCCAACACTAAACCTCACCTTTTTGTCCCCGCTCCCTCCC
CTAGAGATGGGTGCAGAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCA
GGCACCAGGCAGAGACATCACAAAGATTGGTGGGGATCAGTCAGCCTCGGAACCCCTTGGAAGAGCTCAGGCCTAGC
CCTAGGGGTCAAAATGTGGGGCCTGGGCCCCCTGCCCAGACAGGTACCTGTTGGAGCCATGGTAACACGGCCTCCAT
GGCTGAGTAGGGGACTAGGAAGGGTAAAAGTGGGGTTTTGGGGTTTTGCACTCACTCCTGCTGTCTCCTACTTACTG
TGATAGTCCTGGTCCCAGCTCCTGGAAAGCTCTTTGCTCTTAGAAGATCTCCCTTTCCTCCAGCAAAATGTCATCTC
GCCAGGTGCCATGGCTTGTACCTGTAATCACAGCTACTCAGGAGGCTGAAGCAGGAGGATCACTGGAGGCCAGGAGT
TGGAGACCAGCCTGGATGACAGAGGGAGATCCCATCTCTTTAAAAAATAAATAAATAAATAAATAATAAATATC

SEQ ID NO:1491
>T86235 # transcript_30 #len 752 (Includes node 11 - TAA seg 14)
CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGGTGGGGATCAGTCAGCCTCGGAACCCCTTGGAAGAGCTCAGGCCTAGCCCTAGGGGTCAAAAT
GTGGGGCCTGGGCCCCCTGCCCAGACAGGTACCTGTTGGAGCCATGGTAACACGGCCTCCATGGCTGAGTAGGGGAC
TAGGAAGGGTAAAAGTGGGGTTTTGGGGTTTTGCACTCACTCCTGCTGTCTCCTACTTACTGTGATAGTCCTGGTCC
CAGCTCCTGGAAAGCTCTTTGCTCTTAGAAGATCTCCCTTTCCTCCAGCAAAATGTCATCTCGCCAGGTGCCATGGC

TTGTACCTGTAATCACAGCTACTCAGGAGGCTGAAGCAGGAGGATCACTGGAGGCCAGGAGTTGGAGACCAGCCTGG
ATGACAGAGGGAGATCCCATCTCTTTAAAAAATAAATAAATAAATAAATAATAAATATC

SEQ ID NO:1492
>T86235_PEA_13_P8 # trn_8, trn_9 #len 314
MVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQEQ
LEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEVPEPCPPAEPRPLESYCRIEPEIPESSRQEQLEVPEPCPPAE
PGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLRPP
AGQAGLSNLAPRTLALRERLKSCLTAIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQDALCFIPVGSA
APQGSP

SEQ ID NO:1493
>T86235_PEA_13_P9 # trn_10 #len 285
MVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQEQ
LEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEEQLEVPEPCPPAEPGPLQPSTQGQSGPPGPCPRVELGASEPC
TLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLRPPAGQAGLSNLAPRTLALRERLKSCLTAIHC
FHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQDALCFIPVGSAAPQGSP

SEQ ID NO:1494
>T86235_PEA_13_P2 # trn_22 #len 868
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKARHQAE
TSQRLVGISQPRNPLEELRPSPRGQNVGPGPPAQTEAPGTIEFVADPAALATILSGEGVKSCHLGRQPSLAKRVLVR
GSQGGTTQRVQGVRASAYLAPRTPTHRLDPARASCFSRLEGPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELR
RETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHSDEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSP
APVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRRLTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPA
LPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTPSLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSS
LDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQ
EQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEVPEPCPPAEPRPLESYCRIEPEIPESSRQEQLEVPEPCPP
AEPGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLR
PPAGQAGKELAGKECEHKRSSPHCELHTCPAPTPGNLMLPHLPPWPSLALPQEEGRGCTSSPVLLIGLAVGGGGGED
STWWKYRTPDLPLNFPCPSGLSNLAPRTLALRERLKSCLTAIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATL
LEWQDALCFIPVGSAAPQGSP

SEQ ID NO:1495
>T86235_PEA_13_P4 # trn_23 #len 901
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKARHQAE
TSQRLVGISQPRNPLEELRPSPRGQNVGPGPPAQTEAPGTIEFVADPAALATILSGEGVKSCHLGRQPSLAKRVLVR
GSQGGTTQRVQGVRASAYLAPRTPTHRLDPARASCFSRLEGPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELR

RETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHSDEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSP
APVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRRLTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPA
LPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTPSLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSS
LDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQ
EQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEVPEPCPPAEPRPLESYCRIEPEIPESSRQEQLEVPEPCPP
AEPGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLR
PPAGQAGKELAGKECEHKRSSPHCELHTCPAPTPGNLMLPHLPPWPSLALPQEEGRGCTSSPVLLIGLAVGGGGGED
STWWKYRTPDLPLNFPCPSGLSNLAPRTLALRERLKSCLTAIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATL
LEWQDALVRLQPTAQLWLHSEPDGRWGTGTGGHLGFFTERSAGRLGYSARLAEL

SEQ ID NO:1496
>T86235_PEA_13_P5 # trn_24 #len 782
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKARHQAE
TSQRLVGISQPRNPLEELRPSPRGQNVGPGPPAQTEAPGTIEFVADPAALATILSGEGVKSCHLGRQPSLAKRVLVR
GSQGGTTQRVQGVRASAYLAPRTPTHRLDPARASCFSRLEGPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELR
RETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHSDEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSP
APVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRRLTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPA
LPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTPSLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSS
LDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPKPCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQ
EQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEEQLEVPEPCPPAEPGPLQPSTQGQSGPPGPCPRVELGASE
PCTLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLRPPAGQAGLSNLAPRTLALRERLKSCLTAI
HCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQDALVRLQPTAQLWLHSEPDGRWGTGTGGHLGFFTERSA
GRLGYSARLAEL

SEQ ID NO:1497
>T86235_PEA_13_P18 # trn_25 #len 665
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRGVRASAYLAPRTPTHRLDPARASCFSRLE
GPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELRRETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHS
DEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSPAPVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRR
LTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPALPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTP
SLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSSLDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPK
PCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQEQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEEQL
EVPEPCPPAEPGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLESSLPPCCSQWAPATTSLIFSSQHPLCASPP
ICSLQSLRPPAGQAGLSNLAPRTLALRERLKSCLTAIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQD
ALVRLQPTAQLWLHSEPDGRWGTGTGGHLGFFTERSAGRLGYSARLAEL

SEQ ID NO:1498
>T86235_PEA_13_P19 # trn_26 #len 784

MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRGVRASAYLAPRTPTHRLDPARASCFSRLE
GPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELRRETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHS
DEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSPAPVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRR
LTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPALPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTP
SLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSSLDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPK
PCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQEQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEVPE
PCPPAEPRPLESYCRIEPEIPESSRQEQLEVPEPCPPAEPGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLES
SLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLRPPAGQAGKELAGKECEHKRSSPHCELHTCPAPTPGNLMLP
HLPPWPSLALPQEEGRGCTSSPVLLIGLAVGGGGGEDSTWWKYRTPDLPLNFPCPSGLSNLAPRTLALRERLKSCLT
AIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQDALVRLQPTAQLWLHSEPDGRWGTGTGGHLGFFTER
SAGRLGYSARLAEL

SEQ ID NO:1499
>T86235_PEA_13_P20 # trn_27 #len 751
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRGVRASAYLAPRTPTHRLDPARASCFSRLE
GPGPRGRTLCPQRLQALISPSGPSFHPSTRPSFQELRRETAGSSRTSVSQASGLLLETPVQPAFSLPKGEREVVTHS
DEGGVASLGLAQRVPLRENREMSHTRDSHDSHLMPSPAPVAQPLPGHVVPCPSPFGRAQRVPSPGPPTLTSYSVLRR
LTVQPKTRFTPMPSTPRVQQAQWLRGVSPQSCSEDPALPWEQVAVRLFDQESCIRSLEGSGKPPVATPSGPHSNRTP
SLQEVKIQRIGILQQLLRQEVEGLVGGQCVPLNGGSSLDMVELQPLLTEISRTLNATEHNSGTSHLPGLLKHSGLPK
PCLPEECGEPQPCPPAEPGPPEAFCRSEPEIPEPSLQEQLEVPEPYPPAEPRPLESCCRSEPEIPESSRQEQLEVPE
PCPPAEPRPLESYCRIEPEIPESSRQEQLEVPEPCPPAEPGPLQPSTQGQSGPPGPCPRVELGASEPCTLEHRSLES
SLPPCCSQWAPATTSLIFSSQHPLCASPPICSLQSLRPPAGQAGKELAGKECEHKRSSPHCELHTCPAPTPGNLMLP
HLPPWPSLALPQEEGRGCTSSPVLLIGLAVGGGGGEDSTWWKYRTPDLPLNFPCPSGLSNLAPRTLALRERLKSCLT
AIHCFHEARLDDECAFYTSRAPPSGPTRVCTNPVATLLEWQDALCFIPVGSAAPQGSP

SEQ ID NO:1500
>T86235_PEA_13_P21 # trn_29 #len 52
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRVRGA

SEQ ID NO:1501
>T86235_PEA_13_P13 # trn_30 #len 126
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKARHQAE
TSQRLVGISQPRNPLEELRPSPRGQNVGPGPPAQTGTCWSHGNTASMAE

SEQ ID NO:1502
>T86235 # node_6 (TAA seg 9) #len 92
GTAAGAGGGGCCTAATGGGGGAAGACAGTAGTCACACCAGTAATGCACCCCAACACTAAA

CCTCACCTTTTTGTCCCCGCTCCCTCCCCTAG

SEQ ID NO:1503
>T86235 # node_11 (TAA seg 14) #len 111
GTACCTGTTGGAGCCATGGTAACACGGCCTCCATGGCTGAGTAGGGGACTAGGAAGGGTAAAAGTGGGGTTTTGGGG
TTTTGCACTCACTCCTGCTGTCTCCTACTTACTG

SEQ ID NO:1504
>T86235 # node_44 (TAA seg 35) #len 204
CTGTCTCCACTAAAATTTTAAAAATTAGAAAGTGCTCTCTGGAAAAGCTGCCTAACTCTCACTGCTTCTCTGCTGCC
CCTCCTAATGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGGGTGGGACTCAGGCTGGTCT
TTCTCCTGCCCCAGCCTGGCTTGCTTGTGTGCCTTGTTCCTTGGTGACAG

SEQ ID NO:1505
>T86235 # node_32 (TAA seg 37) #len 153
GTGAGTCTGTGTGGCCAACAGCTTTGATGTCTATTGAACAGTGACTGGGCTGAGGAAGAGGGAAAAGAGATGGGGGA
TCAGGAATAGGACAGTGTGGGTAGACTACTGAACGCACATCTTGATGTCACACTGGGGTGCTCTCTCCCACCACAG

SEQ ID NO:1506
>T86235 # node_37 (TAA seg 42) #len 270
GTAAGGAGTTGGCTGGGAAGGAGTGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGCCCCT
ACCCCAGGCAATCTCATGCTTCCACACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCAGGAAGAGGGGAGGGGCTG
CACTTCCAGCCCTGTGCTCCTAATTGGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACAGTACATGGTGGAAGTATA
GGACCCCAGACCTCCCTCTAAATTTTCCATGCCCCTCAG

SEQ ID NO:1507
>T86235 # node_39 (TAA seg 44) #len 165
GTGAGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGGGGGCCA
CCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTGTGACAAGGTCTTC
TCTGTCTCCAG

SEQ ID NO:1508
>Unique aa coded by node_37 (TAA seg 42) [found in T86235_PEA_13_P2 # trn_22,
T86235_PEA_13_P4 # trn_23, T86235_PEA_13_P20 # trn_27, T86235_PEA_13_P19 #
trn_26,
GKELAGKECEHKRSSPHCELHTCPAPTPGNLMLPHLPPWPSLALPQEEGRGCTSSPVLLIGLAVGGGGGEDSTWWKY
RTPDLPLNFPCPS

SEQ ID NO:1509
>Unique aa coded by node_39 (TAA seg 42) [found in T86235_PEA_13_P4 # trn_23, T86235_PEA_13_P5 # trn_24, T86235_PEA_13_P18 # trn_25, T86235_PEA_13_P19 # trn_26]
VRLQPTAQLWLHSEPDGRWGTGTGGHLGFFTERSAGRLGYSARLAEL

SEQ ID NO:1510
>Unique aa coded by node_11 (TAA seg 14) [found in T86235_PEA_13_P13 # trn_30]
GTCWSHGNTASMAE

SEQ ID NO:1511
>Unique aa coded by node_6 (TAA seg 9) [found in T86235_PEA_13_P21 # trn_29]
VRGA

SEQ ID NO:1512
>Oligo from Seg 14 (T86235_0_0_57365)
CATGGTAACACGGCCTCCATGGCTGAGTAGGGGACTAGGAAGGGTAAAAG

SEQ ID NO:1513
>Oligo from seg 35 (T86235_0_0_57371)
TGTACATCTAGGGCCTCTCAGTTAGGGGCTTCAATCCATTCCTCATGAGG

SEQ ID NO:1514
>Oligo from seg 42 (T86235_0_0_57378)
TGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGC

SEQ ID NO:1515
>forward primer:
GCGAAACGCGATTTGTTGTT

SEQ ID NO:1516
>reverse primer:
CATCTGGAGGAGGGAGGGA

SEQ ID NO:1517

>amplicon
GCGAAACGCGATTTGTTGTTTGTGGGTCTGATTTGTGCGTGCGGCTTGGGCTCCTGCGGCTTTTGGCTCGGCCGGGG
GCCTTGGGCAGCGAGGCTGGAGCCGGAAGAGGTGGAGGTGAAGGGCTGCCCGCCACGTCCCTCCCTCCTCCAGATG

SEQ ID NO:1518
>N31842 # transcript_2 #len 1172 (Includes node 8 - TAA seg 3)
CCCTGGCTTTCGACTAGCGTCCGCTGAGCTCCAGGCTGGTGGCGCGTCACTTAGCTGGGGAAGAGGAGATAAAGGCA
GAAAACACCACAGGAAATTGGCTGACAGCAAAGAGCGGAAGGAAGAAGAGGTGCCCCTATACTAAACACCAGACGCT
GGAATTGGAGAAAGAATTTCTGTTCAATATGTATTTGACGCGAGAGCGCCGCCTGGAGATTAGCAAGACCATTAACC
TTACAGACAGACAAGTCAAAATCTGGTTTCAAAATCGCAGAATGAAACTCAAGAAAATGAACCGAGAGAATCGGATC
CGGGAACTGACCTCCAATTTTAATTTCACCTGAGAGCGCGGCCTCTCCTCCTCCCTTCCCGCTCCTTCCTCTCCCCG
CCCCTCCTCCCTTTGTGCCTGGTGATATATTTTTTTTTCCTCCCTGAGTATAAATGCAATGCGACTGCAAAAAAGGC
AAAGACCTCAGACTCTCCTTCCAAGGGACCTGTGGTTCGTGCTGCGAAGATGCTTCCACTTAAAGCATGAGAAATGG
GGTGCCGGGATGTGGGGTGTGGTGTGTGCCCTCATAGATGGGGGTGGGAGTGTGGCTGGTGTGTGTGTCAAACCCTC
ACTCACCCACGCACTCACACACAGCATTCTGTTCTCCATGCAAAGTTAAGATCGAATCCATCCGCTTGTAGGGGAAA
AAAAGGAAAAAAATTAACCAGAGAGGGTCTGTAATCTCGCAGAGCACAGGCAGAATCGTTCCTTCCTTGCTGCATTT
CCTCCTTAGACTAATAGACGTTTTGGAAAGTTCGGCTAGTGTTCGTGTGTTTGTCGTAGCACCCAGAGCCTCCACCA
AACCCTCTCCATGTCTTTACCTCCCAGTCGCTCTAAGAATCTGCTTGAAGTCTCGTATTTGTACTGCTTTCTGCTTT
TCTCCCACCCCTCCTAGCACCCCCACATCCCCCATCTAGTAACATCTCAGAAATTTCATCCAGAGGAACAAAAAAAT
TAAAAATAGAACATAGCAAAGCAAAGACAGAATGCCCCCCCCCCAAATATTGTCCTGTCCCTGTCTGGGAGTTGTGTT
ATTTAAAGATATTCTGTATGTTGTATCTTTTGCATGTAGCTTCCTTAATGGAGAAAAAAAAACCTAATAAATTTCCA
GAATCATAATCCTCAAA

SEQ ID NO:1519
>N31842 # transcript_3 #len 1489 (Includes node 9 - TAA seg 5 ; node 2 - TAA seg 6)
CTCCCAGGCGCGTGCTGGCTGTGGTTTGCTTTCTCAATGCTGGTGTCCTGGGGAGCTGACGTCCCCCAGCTCAGGTC
AGGGGCTTGCAAAAAGCCTAAAATGGCGATCTTGGGCCAGGGACTAGGGAAGGCTGGGGAGATGGGGGGAGTTCTCT
TTACTGCGTTTTCCCAGTTGAAAATTGTTTCCTGCGAAACGCGATTTGTTGTTTGTGGGTCTGATTTGTGCGTGCGG
CTTGGGCTCCTGCGGCTTTTGGCTCGGCCGGGGGCCTTGGGCAGCGAGGCTGGAGCCGGAAGAGGTGGAGGTGAAGG
GCTGCCCGCCACGTCCCTCCCTCCTCCAGATGCCTGGCTTGGATGGCGTTGGAACAGGGCATTTGGAATGTTAGGAG
ATAAAGGCAGAAAACACCACAGGAAATTGGCTGACAGCAAAGAGCGGAAGGAAGAAGAGGTGCCCCTATACTAAACA
CCAGACGCTGGAATTGGAGAAAGAATTTCTGTTCAATATGTATTTGACGCGAGAGCGCCGCCTGGAGATTAGCAAGA
CCATTAACCTTACAGACAGACAAGTCAAAATCTGGTTTCAAAATCGCAGAATGAAACTCAAGAAAATGAACCGAGAG
AATCGGATCCGGGAACTGACCTCCAATTTTAATTTCACCTGAGAGCGCGGCCTCTCCTCCTCCCTTCCCGCTCCTTC
CTCTCCCCGCCCCTCCTCCCTTTGTGCCTGGTGATATATTTTTTTTTCCTCCCTGAGTATAAATGCAATGCGACTGC
AAAAAAGGCAAAGACCTCAGACTCTCCTTCCAAGGGACCTGTGGTTCGTGCTGCGAAGATGCTTCCACTTAAAGCAT

GAGAAATGGGGTGCCGGGATGTGGGGTGTGGTGTGTGCCCTCATAGATGGGGGTGGGAGTGTGGCTGGTGTGTGTGT
CAAACCCTCACTCACCCACGCACTCACACACAGCATTCTGTTCTCCATGCAAAGTTAAGATCGAATCCATCCGCTTG
TAGGGGAAAAAAAGGAAAAAAATTAACCAGAGAGGGTCTGTAATCTCGCAGAGCACAGGCAGAATCGTTCCTTCCTT
GCTGCATTTCCTCCTTAGACTAATAGACGTTTTGGAAAGTTCGGCTAGTGTTCGTGTGTTTGTCGTAGCACCCAGAG
CCTCCACCAAACCCTCTCCATGTCTTTACCTCCCAGTCGCTCTAAGAATCTGCTTGAAGTCTCGTATTTGTACTGCT
TTCTGCTTTTCTCCCACCCCTCCTAGCACCCCCACATCCCCCATCTAGTAACATCTCAGAAATTTCATCCAGAGGAA
CAAAAAAATTAAAAATAGAACATAGCAAAGCAAAGACAGAATGCCCCCCCCCCAAATATTGTCCTGTCCCTGTCTGGG
AGTTGTGTTATTTAAAGATATTCTGTATGTTGTATCTTTTGCATGTAGCTTCCTTAATGGAGAAAAAAAAAACCTAAT
AAATTTCCAGAATCATAATCCTCAAA

SEQ ID NO:1520
>N31842 # transcript_4 #len 1183 (Includes node 9 - TAA seg 5)
CTCCCAGGCGCGTGCTGGCTGTGGTTTGCTTTCTCAATGCTGGTGTCCTGGGGAGCTGACGTCCCCCAGCTCAGAGG
AGATAAAGGCAGAAAACACCACAGGAAATTGGCTGACAGCAAAGAGCGGAAGGAAGAAGAGGTGCCCCTATACTAAA
CACCAGACGCTGGAATTGGAGAAAGAATTTCTGTTCAATATGTATTTGACGCGAGAGCGCCGCCTGGAGATTAGCAA
GACCATTAACCTTACAGACAGACAAGTCAAAATCTGGTTTCAAAATCGCAGAATGAAACTCAAGAAAATGAACCGAG
AGAATCGGATCCGGGAACTGACCTCCAATTTTAATTTCACCTGAGAGCGCGGCCTCTCCTCCTCCCTTCCCGCTCCT
TCCTCTCCCCGCCCCTCCTCCCTTTGTGCCTGGTGATATATTTTTTTTTCCTCCCTGAGTATAAATGCAATGCGACT
GCAAAAAAGGCAAAGACCTCAGACTCTCCTTCCAAGGGACCTGTGGTTCGTGCTGCGAAGATGCTTCCACTTAAAGC
ATGAGAAATGGGGTGCCGGGATGTGGGGTGTGGTGTGTGCCCTCATAGATGGGGGTGGGAGTGTGGCTGGTGTGTGT
GTCAAACCCTCACTCACCCACGCACTCACACACAGCATTCTGTTCTCCATGCAAAGTTAAGATCGAATCCATCCGCT
TGTAGGGGAAAAAAAGGAAAAAAATTAACCAGAGAGGGTCTGTAATCTCGCAGAGCACAGGCAGAATCGTTCCTTCC
TTGCTGCATTTCCTCCTTAGACTAATAGACGTTTTGGAAAGTTCGGCTAGTGTTCGTGTGTTTGTCGTAGCACCCAG
AGCCTCCACCAAACCCTCTCCATGTCTTTACCTCCCAGTCGCTCTAAGAATCTGCTTGAAGTCTCGTATTTGTACTG
CTTTCTGCTTTTCTCCCACCCCTCCTAGCACCCCCACATCCCCCATCTAGTAACATCTCAGAAATTTCATCCAGAGG
AACAAAAAAAATTAAAAATAGAACATAGCAAAGCAAAGACAGAATGCCCCCCCCCCAAATATTGTCCTGTCCCTGTCTG
GGAGTTGTGTTATTTAAAGATATTCTGTATGTTGTATCTTTTGCATGTAGCTTCCTTAATGGAGAAAAAAAAAACCTA
ATAAATTTCCAGAATCATAATCCTCAAA

SEQ ID NO:1521
>N31842_P2 # trn_2 #len 52
MYLTRERRLEISKTINLTDRQVKIWFQNRRMKLKKMNRENRIRELTSNFNFT

SEQ ID NO:1522
>N31842_P4 # trn_4 #len 116
SQARAGCGLLSQCWCPGELTSPSSEEIKAENTTGNWLTAKSGRKKRCPYTKHQTLELEKEFLFNMYLTRERRLEISK
TINLTDRQVKIWFQNRRMKLKKMNRENRIRELTSNFNFT

SEQ ID NO:1523
>Unique aa seq of N31842-P4
SQARAGCGLLSQCWCPGELTSPSS

SEQ ID NO:1524
>N31842 # node_8 (TAA segment 3) #len 63
CCCTGGCTTTCGACTAGCGTCCGCTGAGCTCCAGGCTGGTGGCGCGTCACTTAGCTGGGGAAG

SEQ ID NO:1525
>N31842 # node_9 (TAA segment 5) #len 74
CTCCCAGGCGCGTGCTGGCTGTGGTTTGCTTTCTCAATGCTGGTGTCCTGGGGAGCTGACGTCCCCCAGCTCAG

SEQ ID NO:1526
>N31842 # node_2 (TAA segment 6) #len 306
GTCAGGGGCTTGCAAAAAGCCTAAAATGGCGATCTTGGGCCAGGGACTAGGGAAGGCTGGGGAGATGGGGGGAGTTC
TCTTTACTGCGTTTTCCCAGTTGAAAATTGTTTCCTGCGAAACGCGATTTGTTGTTTGTGGGTCTGATTTGTGCGTG
CGGCTTGGGCTCCTGCGGCTTTTGGCTCGGCCGGGGGCCTTGGGCAGCGAGGCTGGAGCCGGAAGAGGTGGAGGTGA
AGGGCTGCCCGCCACGTCCCTCCCTCCTCCAGATGCCTGGCTTGGATGGCGTTGGAACAGGGCATTTGGAATGTT

SEQ ID NO:1527
>Forward primer:
CTCGCTCCCTTGCTCACAC

SEQ ID NO:1528
>Reverse primer:
AAAGGGAAAGCGGGATGTTT

SEQ ID NO:1529
>Amplicon
CTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACTGACCATTTTGCAAGTGTCAGG
ACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCTTTCCCTTT

SEQ ID NO:1530
>Forward primer:
ACATCCCCCTGGAACGGAT

SEQ ID NO:1531
>Reverse primer:

CAGAAATTAGCAAAGCATTGATGG

SEQ ID NO:1532

>Amplicon

ACATCCCCCTGGAACGGATATCTGTTTGGGGCACTACAATCTATCCTGTAGAACTATGGCCAAATCTCCATCAATGC
TTTGCTAATTTCTG

SEQ ID NO:1533

>T06014 # transcript_3 #len 4096 (Includes node 29 - TAA seg 1 ; node 1 - TAA seg 3)

AGGGGAGTGGGAGGGGTGGGGGTGAGGGGAGAGGCGAGAGGTTTAGCGTGTGGAGCTGCCTGCGCTCCGCCCGGGCT
GTCAGTCCCGGCTCCAGCCGCCGCGAGACCTTCCCGGAGACGCGCGCACACACAGCGCACCCCCTGCACACCGCACA
CCCTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACTGACCATTTTGCAAGTGTCA
GGACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCTTTCCCTTTCCGGGATTTAGTCTGGGAGACGACGACGAGGG
AAGAAGAAACCAGAGGAGTCCTGTCTGGGGGTCCCATCATTATTCGGGATACCCGCCGCCAGCGGCCTGCCTTCGGT
TACCCACATCCCCCTGGAACGGATATCTGTTTGGGGCACTACAATCTATCCTGTAGAACTATGGCCAAATCTCCATC
AATGCTTTGCTAATTTCTGGGACTTAACTCGTAGAATCTACATACAGGGCTGGAATTTATTCAAAATGCATCTGAAG
AAATGACATTTTAAACCGTTTTAAAAAATATCTTGATAAAAAATTCTGTAAAACAGAATTTGATAGGTTTAAAAACA
TGACAGCAGGCTCCCGCTGCGGCCGGGACCTCGCATCCCTGCAACGTGGCCGGGGCTGCATTTTTCATGAGCCTAGG
GTGAACAGGTGCGAAGTGCGCTGGGAGCATCCGGCCAGCGGCCGAGCGCGGGGAACATGGAGAGCGAGCGCGACATG
TACCGCCAGTTCCAGGACTGGTGCCTCAGGACTTACGGGGACTCAGGCAAGACCAAGACGGTGACCCGTAAAAAATA
CGAACGGATCGTCCAGCTCCTCAATGGCTCCGAGTCGAGCTCCACGGACAACGCCAAATTTAAATTCTGGGTCAAAT
CGAAGGGCTTCCAGCTGGGCCAGCCGGACGAGGTCCGCGGGGGAGGCGGCGGCGCCAAGCAAGTGCTCTACGTGCCT
GTCAAGACCACGGATGGCGTAGGGGTAGATGAGAAGCTATCTTTACGACGGGTAGCTGTGGTTGAAGATTTCTTTGA
CATTATTTATTCGATGCATGTGGAAACGGGGCCAAATGGAGAACAAATTCGGAAACACGCTGGACAAAAGAGAACTT
ACAAAGCAATTTCAGAGAGCTATGCCTTCCTACCAAGAGAAGCGGTGACACGATTTCTAATGAGCTGCTCAGAGTGC
CAGAAAAGAATGCATTTAAACCCAGATGGAACAGATCATAAAGATAATGGAAAACCTCCCACTTTGGTGACCAGCAT
GATTGACTACAACATGCCAATTACCATGGCCTACATGAAACACATGAAGCTGCAGCTGCTAAACTCACAGCAAGATG
AGGATGAAAGTTCAATAGAAAGTGATGAATTTGACATGAGTGATTCAACACGGATGTCAGCTGTGAACTCTGATCTT
AGCTCCAATCTTGAAGAAAGAATGCAAAGTCCCCAGAATCTTCATGGCCAGCAAGATGATGATTCTGCTGCAGAGAG
CTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCA
ACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTA
ACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAA
GATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATG
AGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGAC
GATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATGATGAGTCTGCTCCAGC
TGACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGT
ACATCAATGGAAATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCT

TCCAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTC
CAGACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTT
TATTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCACTGAGG
TCTAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTA
AGAGAAATACTTCCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACC
TCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGC
ATAAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGC
TGAAATATTAGTGGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAGAAAAA
AAAAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTT
TTATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAA
ACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAA
ATAACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCT
GCATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAGAAT
TTTTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATT
GCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAA
TCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCT
CCTGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTA
CAATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTGATG
TTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAAC
TTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATT
TATGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTTCAA
GTAATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAG
TGTTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAAT
ATAATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGGATT
TAGAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAA
CAGATTTCTGTGTTA

SEQ ID NO:1534
>T06014 # transcript_5 #len 3081 (Includes node 34 - TAA seg 24)
CTCTTATTTTCCAAAGCTCCAGGTGCTGTTTTCCTTGGGACTCATATACAGAGTCTTTGGTTCAAATGCTATGGTAC
CCAGGCTGAAACCTTGAGAGAAGAACCACCGCCTTGGCCCAGTGAACCTGAAGGCCATACCCACAGACCTAAATACT
GCACCTTCTGTGTAGCAGCTACTCAGGATGGACCAAACTATGTTGGATGATTCTGCTGCAGAGAGCTTTAATGGCAA
TGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCAACAGTGATGGCA
AAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTAACTTCGGAATAC
AGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAAGATGGAACGAGA
GGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATGAGAGTGTAGACC
GAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGACGATTCGGAGAAA
GTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATATGTTTGTCAGGCTGTTTGTAGATGAAAA

CTTGGACCGAATGGTCCCAATCTCTAAGCAGCCCAAAGAAAAGATCCAGGCTATCATTGACTCATGCAGGCGACAAT
TCCCTGAGTATCAAGAGCGTGCCAGAAAACGTATACGTACTTACCTCAAGTCCTGCAGGCGGATGAAAAGAAGTGGT
TTTGAGATGTCTCGACCTATTCCTTCCCACCTTACTTCAGCAGTTGCAGAGAGTATCTTGGCTTCAGCTTGTGAGAG
TGAGAGTAGAAATGCCGCCAAGAGGATGCGTCTGGAGAGACAGCAGGATGAGTCTGCTCCAGCTGACAAACAGTGTA
AACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGTACATCAATGGAAAT
GGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCTTCCAGCAGTGGACC
CACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTCCAGACCCCAGCTGA
GTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTTTATTGCGATCTGCA
GATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCACTGAGGTCTAAATTTGCAGT
TTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTAAGAGAAATACTTCC
ATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACCTCTGAAGTACTTGA
GCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGCATAAAAGGCTATCT
ATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGCTGAAATATTAGTGG
AATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAAGAAAAAAAAAAGCCAAATTT
CTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTTTTATAAAAAGAAAA
CTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAAACTGTTTTGAGAAA
CTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAAATAACACTTGATGC
TGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCTGCATTTCTGCTCTC
ATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAGAATTTTTTTTTTAATAA
TGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATTGCCGCTGCTGTCTC
CAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAATCTGAAAATTTGCA
CTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCTCCTGATATGATTTG
TATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTACAATGAACTGTGCC
CTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTGATGTTTTCTTTGTGCAC
TGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAACTTTTTTCTGTTGTT
TAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATTTATGTCATTCAGTT
TTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTTCAAGTAATGACTGTGAT
TGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAGTGTTTCTAGAATTA
CTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAATATAATGTGAAAAAT
CTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGGATTTAGAGTTTGCTTAT
TTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAACAGATTTCTGTGTT
A

SEQ ID NO:1535
>T06014 # transcript_6 #len 3021 (Includes node 33 - TAA seg 22)
GAGAATTTAGGGGAAGTCAGACTAATCCCTGGGTGCACTTCGAGAGAAAGGTTGGACCTCGCCTGAGGGATGCTGGA
TCTGCAATGATAGAATTGGCCTGCCACAGACATTGGTATTTAAGTTGAAGAGCAGAATTCACAATGATTCTGCTGCA
GAGAGCTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGA

CTCCAACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGC
AGCTAACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGAC
CTCAAGATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAA
AAATGAGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACC
ATGACGATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATATGTTTGTCAGG
CTGTTTGTAGATGAAAACTTGGACCGAATGGTCCCAATCTCTAAGCAGCCCAAAGAAAAGATCCAGGCTATCATTGA
CTCATGCAGGCGACAATTCCCTGAGTATCAAGAGCGTGCCAGAAAACGTATACGTACTTACCTCAAGTCCTGCAGGC
GGATGAAAAGAAGTGGTTTTGAGATGTCTCGACCTATTCCTTCCCACCTTACTTCAGCAGTTGCAGAGAGTATCTTG
GCTTCAGCTTGTGAGAGTGAGAGTAGAAATGCCGCCAAGAGGATGCGTCTGGAGAGACAGCAGGATGAGTCTGCTCC
AGCTGACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGC
TGTACATCAATGGAAATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGAT
GCTTCCAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAA
CTCCAGACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCAT
TTTTATTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCACTG
AGGTCTAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTT
TTAAGAGAAATACTTCCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTG
ACCTCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGG
GGCATAAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCC
AGCTGAAATATTAGTGGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAGAA
AAAAAAAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAA
TTTTTATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAG
AAAACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACA
AAAATAACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAG
CCTGCATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAG
AATTTTTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTG
ATTGCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGA
CAATCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACA
TCTCCTGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAAC
TTACAATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTG
ATGTTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATAC
AACTTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAA
ATTTATGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTT
CAAGTAATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTC
TAGTGTTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTG
AATATAATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGG
ATTTAGAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAAT
AAACAGATTTCTGTGTTA

SEQ ID NO:1536

>T06014 # transcript_7 #len 2889 (Includes node 34 - TAA seg 24)

CTCTTATTTTCCAAAGCTCCAGGTGCTGTTTTCCTTGGGACTCATATACAGAGTCTTTGGTTCAAATGCTATGGTAC
CCAGGCTGAAACCTTGAGAGAAGAACCACCGCCTTGGCCCAGTGAACCTGAAGGCCATACCCACAGACCTAAATACT
GCACCTTCTGTGTAGCAGCTACTCAGGATGGACCAAACTATGTTGGATGATTCTGCTGCAGAGAGCTTTAATGGCAA
TGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCAACAGTGATGGCA
AAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTAACTTCGGAATAC
AGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAAGATGGAACGAGA
GGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATGAGAGTGTAGACC
GAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGACGATTCGGAGAAA
GTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATTCTCGACCTATTCCTTCCCACCTTACTTC
AGCAGTTGCAGAGAGTATCTTGGCTTCAGCTTGTGAGAGTGAGAGTAGAAATGCCGCCAAGAGGATGCGTCTGGAGA
GACAGCAGGATGAGTCTGCTCCAGCTGACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCT
GTTCCAGGCTCACAGGACGTGCTGTACATCAATGGAAATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGG
GGGTCTGCTAAATCTGAATGATGCTTCCAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCT
CAGGATCCTCCAGCAGCTCAAACTCCAGACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCA
GGATATCGAGAATCAGCTGCATTTTTATTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGA
CAGACGACCACCATATTCACTGAGGTCTAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACT
TCTGGTCTTACTGCACAGTCTTTTAAGAGAAATACTTCCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTT
AAGGTGCTTCAAAGGAACTCTGACCTCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTA
GTGTGTCACCATAAATATCAGGGGCATAAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGT
ATAAAACAATAGTTCAAGATCCAGCTGAAATATTAGTGGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTAC
CTGGCAAAAAAAAAAAAAAGAAAAAAAAAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCT
CGTATTTTAAAAAAATATGTAATTTTTATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTA
ACTTTTTCATAGGTCCTAAAAGAAAACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTC
CTCTTTCCAAATTCTTTCTACAAAAATAACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACAT
CAGGAACTACTTCCAAGAGAAGCCTGCATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTT
ATCATAGCAGTTTTCATCCCAGAATTTTTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGT
TTCTTAGAGCTGCTCGTTCCTGATTGCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTG
AATGTTAAGAGGTTGCAAGTGACAATCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAAT
AGTTTCCAAAAAGACCATTACATCTCCTGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAAC
TCTCAGCGGATGCAGCTGCAACTTACAATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTA
TAGTGTGGGATACACATGAGTGATGTTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTA
CTTTAGTTCATGCTTCTTATACAACTTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTG
AATTTGCCTTATTTCATGTAAAATTTATGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTAT
CTACTTGTTACATATAGTTTTTCAAGTAATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGG
ACATTGCTATGCTTTTTTCTTCTAGTGTTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAA
AACTTTTGTGATACTGTTGGTGAATATAATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGC

ACAAACATCTTTTAAGATGTGGATTTAGAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTT
TCTCTGTTTTGTTTGGTTTAATAAACAGATTTCTGTGTTA

SEQ ID NO:1537
>T06014 # transcript_11 #len 4406 (Includes node 29 - TAA seg 1 ; node 1 - TAA seg 3)
AGGGGAGTGGGAGGGGTGGGGGTGAGGGGAGAGGCGAGAGGTTTAGCGTGTGGAGCTGCCTGCGCTCCGCCCGGGCT
GTCAGTCCCGGCTCCAGCCGCCGCGAGACCTTCCCGGAGACGCGCGCACACACAGCGCACCCCCTGCACACCGCACA
CCCTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACTGACCATTTTGCAAGTGTCA
GGACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCTTTCCCTTTCCGGGATTTAGTCTGGGAGACGACGACGAGGG
AAGAAGAAACCAGAGGAGTCCTGTCTGGGGGTCCCATCATTATTCGGGATACCCGCCGCCAGCGGCCTGCCTTCGGT
TACCCACATCCCCCTGGAACGGATATCTGTTTGGGGCACTACAATCTATCCTGTAGAACTATGGCCAAATCTCCATC
AATGCTTTGCTAATTTCTGGGACTTAACTCGTAGAATCTACATACAGGGCTGGAATTTATTCAAAATGCATCTGAAG
AAATGACATTTTAAACCGTTTTAAAAAATATCTTGATAAAAAATTCTGTAAAACAGAATTTGATAGGTTTAAAAACA
TGACAGCAGGCTCCCGCTGCGGCCGGGACCTCGCATCCCTGCAACGTGGCCGGGGCTGCATTTTTCATGAGCCTAGG
GTGAACAGGTGCGAAGTGCGCTGGGAGCATCCGGCCAGCGGCCGAGCGCGGGGAACATGGAGAGCGAGCGCGACATG
TACCGCCAGTTCCAGGACTGGTGCCTCAGGACTTACGGGGACTCAGGCAAGACCAAGACGGTGACCCGTAAAAAATA
CGAACGGATCGTCCAGCTCCTCAATGGCTCCGAGTCGAGCTCCACGGACAACGCCAAATTTAAATTCTGGGTCAAAT
CGAAGGGCTTCCAGCTGGGCCAGCCGGACGAGGTCCGCGGGGGAGGCGGCGGCGCCAAGCAAGTGCTCTACGTGCCT
GTCAAGACCACGGATGGCGTAGGGGTAGATGAGAAGCTATCTTTACGACGGGTAGCTGTGGTTGAAGATTTCTTTGA
CATTATTTATTCGATGCATGTGGAAACGGGGCCAAATGGAGAACAAATTCGGAAACACGCTGGACAAAAGAGAACTT
ACAAAGCAATTTCAGAGAGCTATGCCTTCCTACCAAGAGAAGCGGTGACACGATTTCTAATGAGCTGCTCAGAGTGC
CAGAAAAGAATGCATTTAAACCCAGATGGAACAGATCATAAAGATAATGGAAAACCTCCCACTTTGGTGACCAGCAT
GATTGACTACAACATGCCAATTACCATGGCCTACATGAAACACATGAAGCTGCAGCTGCTAAACTCACAGCAAGATG
AGGATGAAAGTTCAATAGAAAGTGATGAATTTGACATGAGTGATTCAACACGGATGTCAGCTGTGAACTCTGATCTT
AGCTCCAATCTTGAAGAAAGAATGCAAAGTCCCCAGAATCTTCATGGCCAGCAAGATGATGATTCTGCTGCAGAGAG
CTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCA
ACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTA
ACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAA
GATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATG
AGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGAC
GATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATATGTTTGTCAGGCTGTT
TGTAGATGAAAACTTGGACCGAATGGTCCCAATCTCTAAGCAGCCCAAAGAAAAGATCCAGGCTATCATTGACTCAT
GCAGGCGACAATTCCCTGAGTATCAAGAGCGTGCCAGAAAACGTATACGTACTTACCTCAAGTCCTGCAGGCGGATG
AAAAGAAGTGGTTTTGAGATGTCTCGACCTATTCCTTCCCACCTTACTTCAGCAGTTGCAGAGAGTATCTTGGCTTC
AGCTTGTGAGAGTGAGAGTAGAAATGCCGCCAAGAGGATGCGTCTGGAGAGACAGCAGGATGAGTCTGCTCCAGCTG
ACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGTAC
ATCAATGGAAATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCTTC

CAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTCCA
GACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTTTA
TTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCACTGAGGTC
TAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTAAG
AGAAATACTTCCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACCTC
TGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGCAT
AAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGCTG
AAATATTAGTGGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAGAAAAAAA
AAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTTTT
ATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAAAC
TGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAAAT
AACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCTGC
ATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAGAATTT
TTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATTGC
CGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAATC
TGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCTCC
TGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTACA
ATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTGATGTT
TTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAACTT
TTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATTTA
TGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTTCAAGT
AATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAGTG
TTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAATAT
AATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGGATTTA
GAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAACA
GATTTCTGTGTTAAAAA

SEQ ID NO:1538
>T06014 # transcript_12 #len 4214 (Includes node 29 - TAA seg 1 ; node 1 - TAA seg 3)
AGGGGAGTGGGAGGGGTGGGGGTGAGGGGAGAGGCGAGAGGTTTAGCGTGTGGAGCTGCCTGCGCTCCGCCCGGGCT
GTCAGTCCCGGCTCCAGCCGCCGCGAGACCTTCCCGGAGACGCGCGCACACACAGCGCACCCCCTGCACACCGCACA
CCCTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACTGACCATTTTGCAAGTGTCA
GGACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCTTTCCCTTTCCGGGATTTAGTCTGGGAGACGACGACGAGGG
AAGAAGAAACCAGAGGAGTCCTGTCTGGGGGTCCCATCATTATTCGGGATACCCGCCGCCAGCGGCCTGCCTTCGGT
TACCCACATCCCCCTGGAACGGATATCTGTTTGGGGCACTACAATCTATCCTGTAGAACTATGGCCAAATCTCCATC
AATGCTTTGCTAATTTCTGGGACTTAACTCGTAGAATCTACATACAGGGCTGGAATTTATTCAAAATGCATCTGAAG
AAATGACATTTTAAACCGTTTTAAAAAATATCTTGATAAAAAATTCTGTAAAACAGAATTTGATAGGTTTAAAAACA

TGACAGCAGGCTCCCGCTGCGGCCGGGACCTCGCATCCCTGCAACGTGGCCGGGGCTGCATTTTTCATGAGCCTAGG
GTGAACAGGTGCGAAGTGCGCTGGGAGCATCCGGCCAGCGGCCGAGCGCGGGGAACATGGAGAGCGAGCGCGACATG
TACCGCCAGTTCCAGGACTGGTGCCTCAGGACTTACGGGGACTCAGGCAAGACCAAGACGGTGACCCGTAAAAAATA
CGAACGGATCGTCCAGCTCCTCAATGGCTCCGAGTCGAGCTCCACGGACAACGCCAAATTTAAATTCTGGGTCAAAT
CGAAGGGCTTCCAGCTGGGCCAGCCGGACGAGGTCCGCGGGGGAGGCGGCGGCGCCAAGCAAGTGCTCTACGTGCCT
GTCAAGACCACGGATGGCGTAGGGGTAGATGAGAAGCTATCTTTACGACGGGTAGCTGTGGTTGAAGATTTCTTTGA
CATTATTTATTCGATGCATGTGGAAACGGGGCCAAATGGAGAACAAATTCGGAAACACGCTGGACAAAAGAGAACTT
ACAAAGCAATTTCAGAGAGCTATGCCTTCCTACCAAGAGAAGCGGTGACACGATTTCTAATGAGCTGCTCAGAGTGC
CAGAAAAGAATGCATTTAAACCCAGATGGAACAGATCATAAAGATAATGGAAAACCTCCCACTTTGGTGACCAGCAT
GATTGACTACAACATGCCAATTACCATGGCCTACATGAAACACATGAAGCTGCAGCTGCTAAACTCACAGCAAGATG
AGGATGAAAGTTCAATAGAAAGTGATGAATTTGACATGAGTGATTCAACACGGATGTCAGCTGTGAACTCTGATCTT
AGCTCCAATCTTGAAGAAAGAATGCAAAGTCCCCAGAATCTTCATGGCCAGCAAGATGATGATTCTGCTGCAGAGAG
CTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCA
ACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTA
ACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAA
GATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATG
AGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGAC
GATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATTCTCGACCTATTCCTTC
CCACCTTACTTCAGCAGTTGCAGAGAGTATCTTGGCTTCAGCTTGTGAGAGTGAGAGTAGAAATGCCGCCAAGAGGA
TGCGTCTGGAGAGACAGCAGGATGAGTCTGCTCCAGCTGACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTAC
TCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGTACATCAATGGAAATGGGACCTATAGTTACCATAGTTACAG
AGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCTTCCAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAAT
TGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTCCAGACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGA
CAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTTTATTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACA
ACAGAACTGAGACAGACGACCACCATATTCACTGAGGTCTAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCC
AACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTAAGAGAAATACTTCCATTATGCCACATTGTCCTTGATCCGT
AAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACCTCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTG
CTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGCATAAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAA
GAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGCTGAAATATTAGTGGAATTTGCTACTGACTCATTGGACTGA
AAGCTGAAGTACCTGGCAAAAAAAAAAAAAAGAAAAAAAAAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAA
TGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTTTTATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATT
TTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAAACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCC
CTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAAATAACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTT
CAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCTGCATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAA
CACTGCAACTTTATCATAGCAGTTTTCATCCCAGAATTTTTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAA
TCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATTGCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGA
TGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAATCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATT
TCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCTCCTGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAAT

AATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTACAATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTC
TTTCTTATTTTATAGTGTGGGATACACATGAGTGATGTTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATT
TAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAACTTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTT
TGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATTTATGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGT
GATATTTCTTATCTACTTGTTACATATAGTTTTTCAAGTAATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTC
TTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAGTGTTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAA
AAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAATATAATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAA
ATACATAGCTGCACAAACATCTTTTAAGATGTGGATTTAGAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGG
CTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAACAGATTTCTGTGTTAAAAA

SEQ ID NO:1539

>T06014 # transcript_13 #len 2775 (Includes node 34 - TAA seg 24)

CTCTTATTTTCCAAAGCTCCAGGTGCTGTTTTCCTTGGGACTCATATACAGAGTCTTTGGTTCAAATGCTATGGTAC
CCAGGCTGAAACCTTGAGAGAAGAACCACCGCCTTGGCCCAGTGAACCTGAAGGCCATACCCACAGACCTAAATACT
GCACCTTCTGTGTAGCAGCTACTCAGGATGGACCAAACTATGTTGGATGATTCTGCTGCAGAGAGCTTTAATGGCAA
TGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGACTCCAACAGTGATGGCA
AAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGCAGCTAACTTCGGAATAC
AGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGACCTCAAGATGGAACGAGA
GGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAAAAATGAGAGTGTAGACC
GAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACCATGACGATTCGGAGAAA
GTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATGATGAGTCTGCTCCAGCTGACAAACAGTG
TAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGTACATCAATGGAA
ATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCTTCCAGCAGTGGA
CCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTCCAGACCCCAGCT
GAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTTTATTGCGATCTG
CAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCACTGAGGTCTAAATTTGCA
GTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTAAGAGAAATACTT
CCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACCTCTGAAGTACTT
GAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGCATAAAAGGCTAT
CTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGCTGAAATATTAGT
GGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAAGAAAAAAAAAAGCCAAAT
TTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTTTTATAAAAAGAA
AACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAAACTGTTTTGAGA
AACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAAATAACACTTGAT
GCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCTGCATTTCTGCTC
TCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAGAATTTTTTTTTTAAT
AATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATTGCCGCTGCTGTC
TCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAATCTGAAAATTTG

CACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCTCCTGATATGATT
TGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTACAATGAACTGTG
CCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTGATGTTTTCTTTGTGC
ACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAACTTTTTTCTGTTG
TTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATTTATGTCATTCAG
TTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTTCAAGTAATGACTGTG
ATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAGTGTTTCTAGAAT
TACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAATATAATGTGAAAA
ATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGGATTTAGAGTTTGCTT
ATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAACAGATTTCTGTG
TTA

SEQ ID NO:1540
>T06014 # transcript_14 #len 2829 (Includes node 33 - TAA seg 22)
GAGAATTTAGGGGAAGTCAGACTAATCCCTGGGTGCACTTCGAGAGAAAGGTTGGACCTCGCCTGAGGGATGCTGGA
TCTGCAATGATAGAATTGGCCTGCCACAGACATTGGTATTTAAGTTGAAGAGCAGAATTCACAATGATTCTGCTGCA
GAGAGCTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGA
CTCCAACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGC
AGCTAACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGAC
CTCAAGATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAA
AAATGAGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACC
ATGACGATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATTCTCGACCTATT
CCTTCCCACCTTACTTCAGCAGTTGCAGAGAGTATCTTGGCTTCAGCTTGTGAGAGTGAGAGTAGAAATGCCGCCAA
GAGGATGCGTCTGGAGAGACAGCAGGATGAGTCTGCTCCAGCTGACAAACAGTGTAAACCAGAGGCGACCCAGGCCA
CTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGTGCTGTACATCAATGGAAATGGGACCTATAGTTACCATAGT
TACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATGATGCTTCCAGCAGTGGACCCACTGATCTCAGCATGAAGAG
ACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCAAACTCCAGACCCCAGCTGAGTCCAACTGAAATCAATGCCG
TGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGCATTTTTATTGCGATCTGCAGATGAACTGGAAAATCTCATT
TTACAACAGAACTGAGACAGACGACCACCATATTCACTGAGGTCTAAATTTGCAGTTTCCACTAATGACATTTTGAT
TTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTCTTTTAAGAGAAATACTTCCATTATGCCACATTGTCCTTGA
TCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTCTGACCTCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCC
TATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCAGGGGCATAAAAGGCTATCTATTCTTAATTCAAGGATAAAA
CAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGATCCAGCTGAAATATTAGTGGAATTTGCTACTGACTCATTGG
ACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAGAAAAAAAAAAGCCAAATTTCTTGTTGCTACAGGATATAAC
AACAATGAAAAGGATCTCGTATTTTAAAAAAATATGTAATTTTTATAAAAAGAAAACTTGTTTTTCATTCAAACTTG
TCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAAAGAAAACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCA
CATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTACAAAAATAACACTTGATGCTGGAAAAACCCTTGCCTACGT
TCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGAAGCCTGCATTTCTGCTCTCATGCTGATCTCAAAAACCCCA

CTCAACACTGCAACTTTATCATAGCAGTTTTCATCCCAGAATTTTTTTTTTAATAATGACAAGACATGTTGTTGAAA
AAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCCTGATTGCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCAC
CTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGTGACAATCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTT
TCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTACATCTCCTGATATGATTTGTATAATTTTCAGTTCTAGCTA
AAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCAACTTACAATGAACTGTGCCCTCCTATCCCCCATACTTTAC
CCTTCTTTCTTATTTTATAGTGTGGGATACACATGAGTGATGTTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAA
ATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTATACAACTTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAA
TTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTAAAATTTATGTCATTCAGTTTTTGACAATGAGTTTGAGGCA
TCAGTGATATTTCTTATCTACTTGTTACATATAGTTTTTCAAGTAATGACTGTGATTGTGACCGAGTAATGTGCACT
TTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCTTCTAGTGTTTCTAGAATTACTGTTCCTTACAATTATGTAA
ACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGGTGAATATAATGTGAAAAATCTTATTGAAATATGAGTATTT
TGGAAATACATAGCTGCACAAACATCTTTTAAGATGTGGATTTAGAGTTTGCTTATTTAAATGAAAATTCAAAAATT
GAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTAATAAACAGATTTCTGTGTTA

SEQ ID NO:1541

>T06014 # transcript_15 #len 2715 (Includes node 33 - TAA seg 22)

GAGAATTTAGGGGAAGTCAGACTAATCCCTGGGTGCACTTCGAGAGAAAGGTTGGACCTCGCCTGAGGGATGCTGGA
TCTGCAATGATAGAATTGGCCTGCCACAGACATTGGTATTTAAGTTGAAGAGCAGAATTCACAATGATTCTGCTGCA
GAGAGCTTTAATGGCAATGAGACTCTGGGGCACAGTTCAATTGCTTCAGGGGGAACACACAGCAGGGAGATGGGAGA
CTCCAACAGTGATGGCAAAACTGGGCTGGAGCAAGATGAACAGCCACTGAACCTGAGTGACAGTCCCCTCTCTGCGC
AGCTAACTTCGGAATACAGAATAGATGATCACAACAGTAATGGGAAAAACAAGTATAAGAATCTTCTAATTTCTGAC
CTCAAGATGGAACGAGAGGCGAGAGAAAATGGAAGCAAGTCTCCTGCACATAGTTACTCCAGCTATGACTCTGGCAA
AAATGAGAGTGTAGACCGAGGAGCTGAGGACCTCTCACTAAACAGGGGAGATGAGGACGAAGATGACCACGAGGACC
ATGACGATTCGGAGAAAGTTAATGAGACAGACGGCGTTGAAGCCGAGCGGCTGAAAGCTTTTAATGATGAGTCTGCT
CCAGCTGACAAACAGTGTAAACCAGAGGCGACCCAGGCCACTTACTCAACATCAGCTGTTCCAGGCTCACAGGACGT
GCTGTACATCAATGGAAATGGGACCTATAGTTACCATAGTTACAGAGGGCTAGGAGGGGGTCTGCTAAATCTGAATG
ATGCTTCCAGCAGTGGACCCACTGATCTCAGCATGAAGAGACAATTGGCGACTAGCTCAGGATCCTCCAGCAGCTCA
AACTCCAGACCCCAGCTGAGTCCAACTGAAATCAATGCCGTGAGACAGCTTGTTGCAGGATATCGAGAATCAGCTGC
ATTTTTATTGCGATCTGCAGATGAACTGGAAAATCTCATTTTACAACAGAACTGAGACAGACGACCACCATATTCAC
TGAGGTCTAAATTTGCAGTTTCCACTAATGACATTTTGATTTCCCAACAGAGATACTTCTGGTCTTACTGCACAGTC
TTTTAAGAGAAATACTTCCATTATGCCACATTGTCCTTGATCCGTAAGTGATGTGTTAAGGTGCTTCAAAGGAACTC
TGACCTCTGAAGTACTTGAGCTACTTTAGTATGTCCAGCCTATTGCTTTTTGTTTTAGTGTGTCACCATAAATATCA
GGGGCATAAAAGGCTATCTATTCTTAATTCAAGGATAAAACAGAAGAAGCTTGTGGTATAAAACAATAGTTCAAGAT
CCAGCTGAAATATTAGTGGAATTTGCTACTGACTCATTGGACTGAAAGCTGAAGTACCTGGCAAAAAAAAAAAAAAG
AAAAAAAAAAGCCAAATTTCTTGTTGCTACAGGATATAACAACAATGAAAAGGATCTCGTATTTTAAAAAAATATGT
AATTTTTATAAAAAGAAAACTTGTTTTTCATTCAAACTTGTCATTTTTACTTTGGTAACTTTTTCATAGGTCCTAAA
AGAAAACTGTTTTGAGAAACTACTGTAAGTACCTTTTCCACATCCCTTTGCCTTCTCCTCTTTCCAAATTCTTTCTA
CAAAAATAACACTTGATGCTGGAAAAACCCTTGCCTACGTTCTTTCAATCGTCACATCAGGAACTACTTCCAAGAGA

AGCCTGCATTTCTGCTCTCATGCTGATCTCAAAAACCCCACTCAACACTGCAACTTTATCATAGCAGTTTTCATCCC
AGAATTTTTTTTTTAATAATGACAAGACATGTTGTTGAAAAAAAATCACACCTTGGTTTCTTAGAGCTGCTCGTTCC
TGATTGCCGCTGCTGTCTCCAGGCATCCCTCTAGCAGCACCTGGATGTAGATGACTGAATGTTAAGAGGTTGCAAGT
GACAATCTGAAAATTTGCACTCTTGTGTGTAGTTTTCTTTTCATTTCTTTCAGAAATAGTTTCCAAAAAGACCATTA
CATCTCCTGATATGATTTGTATAATTTTCAGTTCTAGCTAAAAATAATGTAAGGAACTCTCAGCGGATGCAGCTGCA
ACTTACAATGAACTGTGCCCTCCTATCCCCCATACTTTACCCTTCTTTCTTATTTTATAGTGTGGGATACACATGAG
TGATGTTTTCTTTGTGCACTGAGACAAGCCTATTTTTTAAATATTTAGGGAGAAGTACTTTAGTTCATGCTTCTTAT
ACAACTTTTTTCTGTTGTTTAGCTTTGGTTGGATTACAAATTCTTTGTGCATTCCTGAATTTGCCTTATTTCATGTA
AAATTTATGTCATTCAGTTTTTGACAATGAGTTTGAGGCATCAGTGATATTTCTTATCTACTTGTTACATATAGTTT
TTCAAGTAATGACTGTGATTGTGACCGAGTAATGTGCACTTTTTCTTGTAACTGTGGACATTGCTATGCTTTTTTCT
TCTAGTGTTTCTAGAATTACTGTTCCTTACAATTATGTAAACAAAAAACAAAAAAAAAACTTTTGTGATACTGTTGG
TGAATATAATGTGAAAAATCTTATTGAAATATGAGTATTTTGGAAATACATAGCTGCACAAACATCTTTTAAGATGT
GGATTTAGAGTTTGCTTATTTAAATGAAAATTCAAAAATTGAGGGCTGGTATAATTTTCTCTGTTTTGTTTGGTTTA
ATAAACAGATTTCTGTGTTA

SEQ ID NO:1542
>T06014_PEA_2_P2 # trn_3 #len 536
MESERDMYRQFQDWCLRTYGDSGKTKTVTRKKYERIVQLLNGSESSSTDNAKFKFWVKSKGFQLGQPDEVRGGGGGA
KQVLYVPVKTTDGVGVDEKLSLRRVAVVEDFFDIIYSMHVETGPNGEQIRKHAGQKRTYKAISESYAFLPREAVTRF
LMSCSECQKRMHLNPDGTDHKDNGKPPTLVTSMIDYNMPITMAYMKHMKLQLLNSQQDEDESSIESDEFDMSDSTRM
SAVNSDLSSNLEERMQSPQNLHGQQDDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSD
SPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDE
DDHEDHDDSEKVNETDGVEAERLKAFNDESAPADKQCKPEATQATYSTSAVPGSQDVLYINGNGTYSYHSYRGLGGG
LLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRESAAFLLRSADELENLILQQN

SEQ ID NO:1543
> T06014_PEA_2_P4 # trn_5 #len 387
MDQTMLDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNG
KNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEA
ERLKAFNMFVRLFVDENLDRMVPISKQPKEKIQAIIDSCRRQFPEYQERARKRIRTYLKSCRRMKRSGFEMSRPIPS
HLTSAVAESILASACESESRNAAKRMRLERQQDESAPADKQCKPEATQATYSTSAVPGSQDVLYINGNGTYSYHSYR
GLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRESAAFLLRSADELENLILQ
QN

SEQ ID NO:1544
>Unique region of T06014_PEA_2_P4 and P6 (coded by Node 34-TAA seg 24)
MDQTML

SEQ ID NO:1545

>T06014_PEA_2_P5 # trn_6 #len 353

MGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYD
SGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEAERLKAFNMFVRLFVDENLDRMVPISKQPKEKIQA
IIDSCRRQFPEYQERARKRIRTYLKSCRRMKRSGFEMSRPIPSHLTSAVAESILASACESESRNAAKRMRLERQQDE
SAPADKQCKPEATQATYSTSAVPGSQDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSS
SSNSRPQLSPTEINAVRQLVAGYRESAAFLLRSADELENLILQQN

SEQ ID NO:1546

>T06014_PEA_2_P6 # trn_7 #len 323

MDQTMLDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNG
KNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEA
ERLKAFNSRPIPSHLTSAVAESILASACESESRNAAKRMRLERQQDESAPADKQCKPEATQATYSTSAVPGSQDVLY
INGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRESAAFL
LRSADELENLILQQN

SEQ ID NO:1547

>T06014_PEA_2_P1 # trn_11 #len 638

MESERDMYRQFQDWCLRTYGDSGKTKTVTRKKYERIVQLLNGSESSSTDNAKFKFWVKSKGFQLGQPDEVRGGGGGA
KQVLYVPVKTTDGVGVDEKLSLRRVAVVEDFFDIIYSMHVETGPNGEQIRKHAGQKRTYKAISESYAFLPREAVTRF
LMSCSECQKRMHLNPDGTDHKDNGKPPTLVTSMIDYNMPITMAYMKHMKLQLLNSQQDEDESSIESDEFDMSDSTRM
SAVNSDLSSNLEERMQSPQNLHGQQDDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSD
SPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDE
DDHEDHDDSEKVNETDGVEAERLKAFNMFVRLFVDENLDRMVPISKQPKEKIQAIIDSCRRQFPEYQERARKRIRTY
LKSCRRMKRSGFEMSRPIPSHLTSAVAESILASACESESRNAAKRMRLERQQDESAPADKQCKPEATQATYSTSAVP
GSQDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGY
RESAAFLLRSADELENLILQQN

SEQ ID NO:1548

>T06014_PEA_2_P7 # trn_12 #len 574

MESERDMYRQFQDWCLRTYGDSGKTKTVTRKKYERIVQLLNGSESSSTDNAKFKFWVKSKGFQLGQPDEVRGGGGGA
KQVLYVPVKTTDGVGVDEKLSLRRVAVVEDFFDIIYSMHVETGPNGEQIRKHAGQKRTYKAISESYAFLPREAVTRF
LMSCSECQKRMHLNPDGTDHKDNGKPPTLVTSMIDYNMPITMAYMKHMKLQLLNSQQDEDESSIESDEFDMSDSTRM
SAVNSDLSSNLEERMQSPQNLHGQQDDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSD
SPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDE
DDHEDHDDSEKVNETDGVEAERLKAFNSRPIPSHLTSAVAESILASACESESRNAAKRMRLERQQDESAPADKQCKP
EATQATYSTSAVPGSQDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSP
TEINAVRQLVAGYRESAAFLLRSADELENLILQQN

SEQ ID NO:1549
>T06014_PEA_2_P8 # trn_13 #len 285
MDQTMLDDSAAESFNGNETLGHSSIASGGTHSREMGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNG
KNKYKNLLISDLKMEREARENGSKSPAHSYSSYDSGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEA
ERLKAFNDESAPADKQCKPEATQATYSTSAVPGSQDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQ
LATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRESAAFLLRSADELENLILQQN

SEQ ID NO:1550
>T06014_PEA_2_P9 # trn_14 #len 289
MGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYD
SGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEAERLKAFNSRPIPSHLTSAVAESILASACESESRN
AAKRMRLERQQDESAPADKQCKPEATQATYSTSAVPGSQDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLS
MKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRESAAFLLRSADELENLILQQN

SEQ ID NO:1551
>T06014_PEA_2_P10 # trn_15 #len 251
MGDSNSDGKTGLEQDEQPLNLSDSPLSAQLTSEYRIDDHNSNGKNKYKNLLISDLKMEREARENGSKSPAHSYSSYD
SGKNESVDRGAEDLSLNRGDEDEDDHEDHDDSEKVNETDGVEAERLKAFNDESAPADKQCKPEATQATYSTSAVPGS
QDVLYINGNGTYSYHSYRGLGGGLLNLNDASSSGPTDLSMKRQLATSSGSSSSSNSRPQLSPTEINAVRQLVAGYRE
SAAFLLRSADELENLILQQN

SEQ ID NO:1552
>T06014 # node_29 (TAA seg 1) #len 314
AGGGGAGTGGGAGGGGTGGGGGTGAGGGGAGAGGCGAGAGGTTTAGCGTGTGGAGCTGCCTGCGCTCCGCCCGGGCT
GTCAGTCCCGGCTCCAGCCGCCGCGAGACCTTCCCGGAGACGCGCGCACACACAGCGCACCCCCTGCACACCGCACA
CCCTCGCTCCCTTGCTCACACACACGCACACACTCAGCCTGGCCGAGCAGGAGCCACTGACCATTTTGCAAGTGTCA
GGACCAGCTACAGCGCGGTGGGCGCAAACATCCCGCTTTCCCTTTCCGGGATTTAGTCTGGGAGACGACGACGAGGG
AAGAAG

SEQ ID NO:1553
>T06014 # node_1 (TAA seg 3) #len 308
AAACCAGAGGAGTCCTGTCTGGGGGTCCCATCATTATTCGGGATACCCGCCGCCAGCGGCCTGCCTTCGGTTACCCA
CATCCCCCTGGAACGGATATCTGTTTGGGGCACTACAATCTATCCTGTAGAACTATGGCCAAATCTCCATCAATGCT
TTGCTAATTTCTGGGACTTAACTCGTAGAATCTACATACAGGGCTGGAATTTATTCAAAATGCATCTGAAGAAATGA
CATTTTAAACCGTTTTAAAAAATATCTTGATAAAAAATTCTGTAAAACAGAATTTGATAGGTTTAAAAACATGACAG

SEQ ID NO:1554

>T06014 # node_33 (TAA segment 22) #len 140
GAGAATTTAGGGGAAGTCAGACTAATCCCTGGGTGCACTTCGAGAGAAAGGTTGGACCTCGCCTGAGGGATGCTGGA
TCTGCAATGATAGAATTGGCCTGCCACAGACATTGGTATTTAAGTTGAAGAGCAGAATTCACA

SEQ ID NO:1555
>T06014 # node_34 (TAA segment 24) #len 200
CTCTTATTTTCCAAAGCTCCAGGTGCTGTTTTCCTTGGGACTCATATACAGAGTCTTTGGTTCAAATGCTATGGTAC
CCAGGCTGAAACCTTGAGAGAAGAACCACCGCCTTGGCCCAGTGAACCTGAAGGCCATACCCACAGACCTAAATACT
GCACCTTCTGTGTAGCAGCTACTCAGGATGGACCAAACTATGTTGG

SEQ ID NO:1556
>forward primer:
GGTTCGGATGGACTACACTTTGTC

SEQ ID NO:1557
>Reverse primer:
CCACGTACTTCTGGGTGATGTC

SEQ ID NO:1558
>Amplicon:
GGTTCGGATGGACTACACTTTGTCCGTACCCACCACGTAGCAGAGAAAACCACCTTGTATGACATGGACATTGACAT
CACCCAGAAGTACGTGG

SEQ ID NO:1559
>AA281370 # transcript_0 #len 4783 (Includes node 31 - TAA seg 10)
CGGCGTGACGATGGCGGCCGTAGGGTCCGGAGGCTATGCGCGGAACGATGCAGGGGAGAAGCTGCCCTCTGTCATGG
CGGGAGTTCCGGCGCGGAGGGGCCAGTCCTCCCCGCCCCCCGCCCCACCAATCTGCCTACGGCGGCGGACGCGACTC
TCGACGGCCTCCGAGGAGACGGTGCAGAACCGGGTGTCACTCGAGAAGGTGCTTGGCATCACAGCCCAGAACAGCAG
TGGCCTAACCTGTGACCCCGGCACAGGCCATGTGGCCTACCTGGCAGGCTGTGTGGTGGTGATTTTGGACCCCAAGG
AGAACAAGCAGCAGCACATCTTTAACACCGCCAGGAAGTCTCTCAGTGCTCTGGCCTTCTCCCCTGATGGGAAGTAC
ATAGTGACAGGGGAGAATGGGCATAGGCCTGCTGTGCGCATCTGGGATGTGGAGGAGAAGAATCAGGTGGCGGAGAT
GCTAGGCCACAAGTATGGTGTGGCGTGTGTGGCCTTCTCACCCAATATGAAGCACATCGTGTCCATGGGCTACCAAC
ATGACATGGTGCTCAACGTCTGGGACTGGAAGAAAGACATCGTAGTGGCCTCCAACAAGGTATCTTGTAGAGTCATT
GCCCTCTCCTTCTCAGAGGACAGCAGCTATTTTGTCACTGTTGGGAACCGCCATGTGAGGTTCTGGTTCTTGGAAGT
CTCCACTGAGACAAAGGTGACGAGCACAGTGCCCCTTGTAGGGCGCTCGGGCATCCTGGGCGAGCTGCACAACAACA
TCTTCTGTGGTGTGGCCTGCGGTCGGGGCCGGATGGCGGGCAGTACCTTCTGTGTGTCCTACTCGGGCCTCCTCTGC
CAGTTCAATGAGAAGAGGGTGCTGGAGAAGTGGATCAACCTGAAGGTCTCCCTGTCTTCCTGCCTCTGTGTCAGCCA
GGAGCTCATCTTCTGTGGCTGCACAGATGGGATAGTCCGCATCTTCCAGGCCCATAGCCTGCACTACCTCGCCAACC

TGCCCAAGCCACACTACCTTGGGGTAGACGTGGCACAGGGCCTGGAGCCCAGGAAGGCGGAAGCAGTCTACCCAGAT
ACAGTGGCACTGACCTTCGACCCCATCCACCAGTGGCTGTCCTGCGTGTATAAGGACCACAGCATCTACATCTGGGA
TGTCAAGGACATCAACAGAGTGGGCAAGGTGTGGTCAGAGCTCTTCCACAGCTCCTACGTTTGGAACGTGGAGGTGT
ATCCTGAGTTTGAAGACCAGAGAGCTTGTTTGCCATCAGGATCCTTTCTGACTTGTTCTTCAGACAACACCATTCGC
TTCTGGAACTTGGACAGCAGCCCTGATTCTCACTGGCAGAAAAACATCTTCAGCAATTTGGCTCCTGCTCTGCAGAT
GTGTGGACGTGGCCACTCCCGGCAGCCCAACACACCCAGCCCAGGTGAAATTGCTTCCTAAGTCTTGAGATGGAATT
CCCTGGGAGAGCACCTGATGAGCCAGCCCATTCAGCTGCCACACAACCCTGCTGAAGGTCGTGTACGTGGAGAATGA
CATCCAGCACCTGCAGGACATGTCACACTTCCCAGACCGGGGGAGCGAGAATGGGACACCCATGGACGTGAAAGCCG
GGGTGCGGGTCATGCAGGTCAGTCCTGACGGCCAGCATTTGGCTTCAGGCGACCGAAGTGGAAATCTGAGGATCCAC
GAGCTGCACTTCATGGACGAGCTGGTCAAGGTGGAGGCCCATGATGCTGAGGTGCTGTGCCTGGAGTACTCCAAGCC
AGAGACGGGGCTGACCTTGCTGGCCTCAGCCAGTCGGGACCGGCTGATCCATGTGCTGAACGTGGAGAAGAACTACA
ACCTGGAGCAGACGCTGGATGACCACTCCTCCTCCATCACCGCCATCAAGTTCGCTGGCAACAGAGACATCCAGATG
ATCAGCTGTGGGGCTGACAAGAGCATCTACTTTCGCAGTGCCCAGCAGGGTTCGGATGGACTACACTTTGTCCGTAC
CCACCACGTAGCAGAGAAAACCACCTTGTATGACATGGACATTGACATCACCCAGAAGTACGTGGCCGTGGCCTGCC
AGGACCGCAATGTGAGAGTCTACAACACTGTGAACGGGAAGCAGAAGAAGTGCTACAAGGGCTCCCAGGGTGACGAA
GGGTCCTTGCTGAAGGTCCATGTGGACCCCTCAGGCACCTTCCTGGCCACCAGCTGCTCTGACAAAAGCATCTCAGT
GATTGACTTTTACTCGGGCGAGTGCATTGCCAAGATGTTTGGCCATTCAGAAATTATTACCAGCATGAAGTTCACCT
ATGACTGTCATCACTTGATCACAGTATCTGGAGACAGCTGCGTGTTCATCTGGCACCTGGGCCCGGAGATCACCAAC
TGCATGAAGCAGCACTTGCTGGAGATTGACCACCGGCAGCAGCAGCAGCACACAAATGACAAGAAGCGGAGTGGCCA
CCCCAGGCAGGATACGTATGTGTCCACACCTAGTGAGATTCACTCCCTGAGCCCTGGAGAGCAAACAGAGGATGATC
TGGAGGAAGAGTGTGAGCCAGAAGAGATGCTGAAGACACCATCCAAAGATAGCTTGGATCCAGATCCTCGTTGCCTG
CTAACCAACGGCAAGCTGCCACTGTGGGCAAAGCGGCTGCTAGGGGACGATGATGTGGCAGATGGCTCGGCCTTCCA
CGCCAAGCGCAGCTACCAGCCCCACGGCCGCTGGGCAGAGCGGGCCGGCCAAGAGCCCCTCAAGACCATCCTGGATG
CCCAGGACCTGGATTGCTACTTTACCCCCATGAAGCCCGAGAGTCTGGAGAACTCCATTCTGGATTCACTGGAGCCA
CAGAGCCTGGCCAGCCTGCTGAGTGAGTCAGAGAGTCCCCAGGAAGCTGGCCGCGGGCACCCCTCCTTCCTGCCCCA
GCAGAAGGAATCATCTGAGGCCAGTGAGCTCATCCTCTACTCTCTGGAGGCAGAAGTGACAGTCACAGGGACAGACA
GCCAGTATTGCAGGAAGGAGGTGGAGGCCGGGCCTGGAGACCAGCAGGGCGACTCCTACCTCAGGGTGTCCTCCGAC
AGCCCAAAGGACCAGAGCCCGCCTGAGGACTCGGGGGAGTCAGAGGCCGACCTGGAGTGCAGCTTCGCAGCCATCCA
CTCCCCAGCTCCGCCTCCTGACCCTGCCCCTCGGTTTGCCACGTCGCTGCCCCATTTCCCAGGATGCGCAGGTCCCA
CAGAAGATGAGCTGTCCCTGCCCGAGGGACCCAGCGTCCCCAGCAGCTCCCTACCCCAGACTCCGGAGCAGGAGAAG
TTCCTCCGCCACCACTTTGAGACACTGACTGAGTCCCCCTGCAGAGCTCTGGGAGACGTGGAGGCCTCTGAAGCTGA
AGACCACTTCTTCAACCCACGCCTGAGTATCTCCACGCAGTTCCTCTCAAGCCTCCAGAAGGCATCCAGGTTCACCC
ATACCTTCCCTCCCCGGGCAACCCAGTGCCTTGTGAAGTCTCCAGAGGTCAAGCTCATGGACCGAGGCGGAAGCCAG
CCCAGAGCAGGTACTGGCTACGCCTCCCCAGACAGGACCCACGTCCTCGCTGCAGGGAAGGCTGAAGAGACCCTGGA
GGCCTGGCGCCCACCACCTCCCTGCCTTACGAGCCTGGCGTCCTGTGTCCCTGCTTCCTCCGTGCTGCCCACAGACA
GGAATCTCCCAACGCCCACATCTGCACCCACCCCAGGCCTGGCTCAGGGTGTCCATGCCCCCTCCACCTGTTCCTAC
ATGGAGGCCACTGCCAGCTCCCGTGCCAGGATATCACGCAGCATCTCCCTCGGTGACAGTGAGGGCCCTATCGTGGC
CACACTGGCCCAGCCCCTCCGTAGGCCATCGTCCGTTGGGGAGCTGGCCTCCTTGGGCCAGGAGCTTCAGGCCATCA

CCACCGCGACAACACCCAGTTTGGACAGTGAGGGCCAAGAGCCTGCCCTGCGTTCCTGGGGCAACCACGAGGCCCGG
GCCAACCTGAGACTGACCCTGTCAAGTGCCTGTGATGGGCTCCTGCAGCCCCCCGTGGATACCCAGCCTGGCGTCAC
CGTCCCTGCAGTGAGCTTCCCAGCCCCTAGCCCTGTGGAAGAGAGCGCCCTGAGGCTCCACGGCTCTGCCTTTCGCC
CAAGTCTCCCAGCTCCTGAGTCCCCTGGCCTTCCTGCCCACCCCAGTAACCCCCAGCTTCCAGAGGCCCGGCCTGGC
ATCCCTGGCGGCACTGCCTCCCTCCTGGAGCCCACCTCCGGTGCACTTGGTCTGTTCCAGGGCAGCCCTGCCCGCTG
GAGTGAGCCCTGGGTGCCGGTTGAAGCCCTGCCCCCATCTCCCCTTGAGCTGAGCAGGGTGGGGAACATCTTGCACA
GGCTGCAGACCACCTTCCAAGAAGCCCTCGACCTTTACCGTGTGTTGGTCTCCAGTGGCCAGGTGGACACCGGGCAG
CAGCAGGCACGGACTGAGCTGGTCTCCACCTTCCTGTGGATCCACAGCCAGCTGGAGGCTGAATGCCTGGTGGGGAC
TAGTGTGGCCCCAGCCCAGGCTCTGCCCAGCCCAGGACCCCCGTCCCCACCGACGCTGTACCCCCTGGCCAGCCCAG
ACCTGCAGGCCCTGCTGGAACACTACTCGGAGCTGCTGGTGCAGGCCGTGCGGAGGAAGGCACGGGGGCACTGAGGG
CGCAGCCCCTCCACCGCAGCCCTGCTGCTTCTGAGGACTTAGGTATTTTAAGCGAATAAACTGACAGCTTTGAGGAA
TGGTTCCTG

SEQ ID NO:1560

>AA281370 # transcript_1 #len 4656 (Includes node 31 - TAA seg 10)

CGGCGTGACGATGGCGGCCGTAGGGTCCGGAGGCTATGCGCGGAACGATGCAGGGGAGAAGCTGCCCTCTGTCATGG
CGGGAGTTCCGGCGCGGAGGGGCCAGTCCTCCCCGCCCCCCGCCCCACCAATCTGCCTACGGCGGCGGACGCGACTC
TCGACGGCCTCCGAGGAGACGGTGCAGAACCGGGTGTCACTCGAGAAGGTGCTTGGCATCACAGCCCAGAACAGCAG
TGGCCTAACCTGTGACCCCGGCACAGGCCATGTGGCCTACCTGGCAGGCTGTGTGGTGGTGATTTTGGACCCCAAGG
AGAACAAGCAGCAGCACATCTTTAACACCGCCAGGAAGTCTCTCAGTGCTCTGGCCTTCTCCCCTGATGGGAAGTAC
ATAGTGACAGGGGAGAATGGGCATAGGCCTGCTGTGCGCATCTGGGATGTGGAGGAGAAGAATCAGGTGGCGGAGAT
GCTAGGCCACAAGTATGGTGTGGCGTGTGTGGCCTTCTCACCCAATATGAAGCACATCGTGTCCATGGGCTACCAAC
ATGACATGGTGCTCAACGTCTGGGACTGGAAGAAAGACATCGTAGTGGCCTCCAACAAGGTATCTTGTAGAGTCATT
GCCCTCTCCTTCTCAGAGGACAGCAGCTATTTTGTCACTGTTGGGAACCGCCATGTGAGGTTCTGGTTCTTGGAAGT
CTCCACTGAGACAAAGGTGACGAGCACAGTGCCCCTTGTAGGGCGCTCGGGCATCCTGGGCGAGCTGCACAACAACA
TCTTCTGTGGTGTGGCCTGCGGTCGGGGCCGGATGGCGGGCAGTACCTTCTGTGTGTCCTACTCGGGCCTCCTCTGC
CAGTTCAATGAGAAGAGGGTGCTGGAGAAGTGGATCAACCTGAAGGTCTCCCTGTCTTCCTGCCTCTGTGTCAGCCA
GGAGCTCATCTTCTGTGGCTGCACAGATGGGATAGTCCGCATCTTCCAGGCCCATAGCCTGCACTACCTCGCCAACC
TGCCCAAGCCACACTACCTTGGGGTAGACGTGGCACAGGGCCTGGAGCCCAGCTTCCTCTTCCACAGGAAGGCGGAA
GCAGTCTACCCAGATACAGTGGCACTGACCTTCGACCCCATCCACCAGTGGCTGTCCTGCGTGTATAAGGACCACAG
CATCTACATCTGGGATGTCAAGGACATCAACAGAGTGGGCAAGGTGTGGTCAGAGCTCTTCCACAGCTCCTACGTTT
GGAACGTGGAGGTGTATCCTGAGTTTGAAGACCAGAGAGCTTGTTTGCCATCAGGATCCTTTCTGACTTGTTCTTCA
GACAACACCATTCGCTTCTGGAACTTGGACAGCAGCCCTGATTCTCACTGGCAGAAAAACATCTTCAGCAATACCCT
GCTGAAGGTCGTGTACGTGGAGAATGACATCCAGCACCTGCAGGACATGTCACACTTCCCAGACCGGGGGAGCGAGA
ATGGGACACCCATGGACGTGAAAGCCGGGGTGCGGGTCATGCAGGTCAGTCCTGACGGCCAGCATTTGGCTTCAGGC
GACCGAAGTGGAAATCTGAGGATCCACGAGCTGCACTTCATGGACGAGCTGGTCAAGGTGGAGGCCCATGATGCTGA
GGTGCTGTGCCTGGAGTACTCCAAGCCAGAGACGGGGCTGACCTTGCTGGCCTCAGCCAGTCGGGACCGGCTGATCC
ATGTGCTGAACGTGGAGAAGAACTACAACCTGGAGCAGACGCTGGATGACCACTCCTCCTCCATCACCGCCATCAAG

TTCGCTGGCAACAGAGACATCCAGATGATCAGCTGTGGGGCTGACAAGAGCATCTACTTTCGCAGTGCCCAGCAGGG
TTCGGATGGACTACACTTTGTCCGTACCCACCACGTAGCAGAGAAAACCACCTTGTATGACATGGACATTGACATCA
CCCAGAAGTACGTGGCCGTGGCCTGCCAGGACCGCAATGTGAGAGTCTACAACACTGTGAACGGGAAGCAGAAGAAG
TGCTACAAGGGCTCCCAGGGTGACGAAGGGTCCTTGCTGAAGGTCCATGTGGACCCCTCAGGCACCTTCCTGGCCAC
CAGCTGCTCTGACAAAAGCATCTCAGTGATTGACTTTTACTCGGGCGAGTGCATTGCCAAGATGTTTGGCCATTCAG
AAATTATTACCAGCATGAAGTTCACCTATGACTGTCATCACTTGATCACAGTATCTGGAGACAGCTGCGTGTTCATC
TGGCACCTGGGCCCGGAGATCACCAACTGCATGAAGCAGCACTTGCTGGAGATTGACCACCGGCAGCAGCAGCAGCA
CACAAATGACAAGAAGCGGAGTGGCCACCCCAGGCAGGATACGTATGTGTCCACACCTAGTGAGATTCACTCCCTGA
GCCCTGGAGAGCAAACAGAGGATGATCTGGAGGAAGAGTGTGAGCCAGAAGAGATGCTGAAGACACCATCCAAAGAT
AGCTTGGATCCAGATCCTCGTTGCCTGCTAACCAACGGCAAGCTGCCACTGTGGGCAAAGCGGCTGCTAGGGGACGA
TGATGTGGCAGATGGCTCGGCCTTCCACGCCAAGCGCAGCTACCAGCCCCACGGCCGCTGGGCAGAGCGGGCCGGCC
AAGAGCCCCTCAAGACCATCCTGGATGCCCAGGACCTGGATTGCTACTTTACCCCCATGAAGCCCGAGAGTCTGGAG
AACTCCATTCTGGATTCACTGGAGCCACAGAGCCTGGCCAGCCTGCTGAGTGAGTCAGAGAGTCCCCAGGAAGCTGG
CCGCGGGCACCCCTCCTTCCTGCCCCAGCAGAAGGAATCATCTGAGGCCAGTGAGCTCATCCTCTACTCTCTGGAGG
CAGAAGTGACAGTCACAGGGACAGACAGCCAGTATTGCAGGAAGGAGGTGGAGGCCGGGCCTGGAGACCAGCAGGGC
GACTCCTACCTCAGGGTGTCCTCCGACAGCCCAAAGGACCAGAGCCCGCCTGAGGACTCGGGGGAGTCAGAGGCCGA
CCTGGAGTGCAGCTTCGCAGCCATCCACTCCCCAGCTCCGCCTCCTGACCCTGCCCCTCGGTTTGCCACGTCGCTGC
CCCATTTCCCAGGATGCGCAGGTCCCACAGAAGATGAGCTGTCCCTGCCCGAGGGACCCAGCGTCCCCAGCAGCTCC
CTACCCCAGACTCCGGAGCAGGAGAAGTTCCTCCGCCACCACTTTGAGACACTGACTGAGTCCCCCTGCAGAGCTCT
GGGAGACGTGGAGGCCTCTGAAGCTGAAGACCACTTCTTCAACCCACGCCTGAGTATCTCCACGCAGTTCCTCTCAA
GCCTCCAGAAGGCATCCAGGTTCACCCATACCTTCCCTCCCCGGGCAACCCAGTGCCTTGTGAAGTCTCCAGAGGTC
AAGCTCATGGACCGAGGCGGAAGCCAGCCCAGAGCAGGTACTGGCTACGCCTCCCCAGACAGGACCCACGTCCTCGC
TGCAGGGAAGGCTGAAGAGACCCTGGAGGCCTGGCGCCCACCACCTCCCTGCCTTACGAGCCTGGCGTCCTGTGTCC
CTGCTTCCTCCGTGCTGCCCACAGACAGGAATCTCCCAACGCCCACATCTGCACCCACCCCAGGCCTGGCTCAGGGT
GTCCATGCCCCCTCCACCTGTTCCTACATGGAGGCCACTGCCAGCTCCCGTGCCAGGATATCACGCAGCATCTCCCT
CGGTGACAGTGAGGGCCCTATCGTGGCCACACTGGCCCAGCCCCTCCGTAGGCCATCGTCCGTTGGGGAGCTGGCCT
CCTTGGGCCAGGAGCTTCAGGCCATCACCACCGCGACAACACCCAGTTTGGACAGTGAGGGCCAAGAGCCTGCCCTG
CGTTCCTGGGGCAACCACGAGGCCCGGGCCAACCTGAGACTGACCCTGTCAAGTGCCTGTGATGGGCTCCTGCAGCC
CCCCGTGGATACCCAGCCTGGCGTCACCGTCCCTGCAGTGAGCTTCCCAGCCCCTAGCCCTGTGGAAGAGAGCGCCC
TGAGGCTCCACGGCTCTGCCTTTCGCCCAAGTCTCCCAGCTCCTGAGTCCCCTGGCCTTCCTGCCCACCCCAGTAAC
CCCCAGCTTCCAGAGGCCCGGCCTGGCATCCCTGGCGGCACTGCCTCCCTCCTGGAGCCCACCTCCGGTGCACTTGG
TCTGTTCCAGGGCAGCCCTGCCCGCTGGAGTGAGCCCTGGGTGCCGGTTGAAGCCCTGCCCCCATCTCCCCTTGAGC
TGAGCAGGGTGGGGAACATCTTGCACAGGCTGCAGACCACCTTCCAAGAAGCCCTCGACCTTTACCGTGTGTTGGTC
TCCAGTGGCCAGGTGGACACCGGGCAGCAGCAGGCACGGACTGAGCTGGTCTCCACCTTCCTGTGGATCCACAGCCA
GCTGGAGGCTGAATGCCTGGTGGGGACTAGTGTGGCCCCAGCCCAGGCTCTGCCCAGCCCAGGACCCCCGTCCCCAC
CGACGCTGTACCCCCTGGCCAGCCCAGACCTGCAGGCCCTGCTGGAACACTACTCGGAGCTGCTGGTGCAGGCCGTG
CGGAGGAAGGCACGGGGGCACTGAGGGCGCAGCCCCTCCACCGCAGCCCTGCTGCTTCTGAGGACTTAGGTATTTTA
AGCGAATAAACTGACAGCTTTGAGGAATGGTTCCTG

SEQ ID NO:1561

>AA281370 # transcript_4 #len 4798 (Includes node 31 - TAA seg 10; node 56)

CGGCGTGACGATGGCGGCCGTAGGGTCCGGAGGCTATGCGCGGAACGATGCAGGGGAGAAGCTGCCCTCTGTCATGG

CGGGAGTTCCGGCGCGGAGGGGCCAGTCCTCCCCGCCCCCCGCCCCACCAATCTGCCTACGGCGGCGGACGCGACTC

TCGACGGCCTCCGAGGAGACGGTGCAGAACCGGGTGTCACTCGAGAAGGTGCTTGGCATCACAGCCCAGAACAGCAG

TGGCCTAACCTGTGACCCCGGCACAGGCCATGTGGCCTACCTGGCAGGCTGTGTGGTGGTGATTTTGGACCCCAAGG

AGAACAAGCAGCAGCACATCTTTAACACCGCCAGGAAGTCTCTCAGTGCTCTGGCCTTCTCCCCTGATGGGAAGTAC

ATAGTGACAGGGGAGAATGGGCATAGGCCTGCTGTGCGCATCTGGGATGTGGAGGAGAAGAATCAGGTGGCGGAGAT

GCTAGGCCACAAGTATGGTGTGGCGTGTGTGGCCTTCTCACCCAATATGAAGCACATCGTGTCCATGGGCTACCAAC

ATGACATGGTGCTCAACGTCTGGGACTGGAAGAAAGACATCGTAGTGGCCTCCAACAAGGTATCTTGTAGAGTCATT

GCCCTCTCCTTCTCAGAGGACAGCAGCTATTTTGTCACTGTTGGGAACCGCCATGTGAGGTTCTGGTTCTTGGAAGT

CTCCACTGAGACAAAGGTGACGAGCACAGTGCCCCTTGTAGGGCGCTCGGGCATCCTGGGCGAGCTGCACAACAACA

TCTTCTGTGGTGTGGCCTGCGGTCGGGGCCGGATGGCGGGCAGTACCTTCTGTGTGTCCTACTCGGGCCTCCTCTGC

CAGTTCAATGAGAAGAGGGTGCTGGAGAAGTGGATCAACCTGAAGGTCTCCCTGTCTTCCTGCCTCTGTGTCAGCCA

GGAGCTCATCTTCTGTGGCTGCACAGATGGGATAGTCCGCATCTTCCAGGCCCATAGCCTGCACTACCTCGCCAACC

TGCCCAAGCCACACTACCTTGGGGTAGACGTGGCACAGGGCCTGGAGCCCAGGAAGGCGGAAGCAGTCTACCCAGAT

ACAGTGGCACTGACCTTCGACCCCATCCACCAGTGGCTGTCCTGCGTGTATAAGGACCACAGCATCTACATCTGGGA

TGTCAAGGACATCAACAGAGTGGGCAAGGTGTGGTCAGAGCTCTTCCACAGCTCCTACGTTTGGAACGTGGAGGTGT

ATCCTGAGTTTGAAGACCAGAGAGCTTGTTTGCCATCAGGATCCTTTCTGACTTGTTCTTCAGACAACACCATTCGC

TTCTGGAACTTGGACAGCAGCCCTGATTCTCACTGGCAGAAAAACATCTTCAGCAATTTGGCTCCTGCTCTGCAGAT

GTGTGGACGTGGCCACTCCCGGCAGCCCAACACACCCAGCCCAGGTGAAATTGCTTCCTAAGTCTTGAGATGGAATT

CCCTGGGAGAGCACCTGATGAGCCAGCCCATTCAGCTGCCACACAACCCTGCTGAAGGTCGTGTACGTGGAGAATGA

CATCCAGCACCTGCAGGACATGTCACACTTCCCAGACCGGGGGAGCGAGAATGGGACACCCATGGACGTGAAAGCCG

GGGTGCGGGTCATGCAGGTCAGTCCTGACGGCCAGCATTTGGCTTCAGGCGACCGAAGTGGAAATCTGAGGATCCAC

GAGCTGCACTTCATGGACGAGCTGGTCAAGGTGGAGGCCCATGATGCTGAGGTGCTGTGCCTGGAGTACTCCAAGCC

AGAGACGGGGCTGACCTTGCTGGCCTCAGCCAGTCGGGACCGGCTGATCCATGTGCTGAACGTGGAGAAGAACTACA

ACCTGGAGCAGACGCTGGATGACCACTCCTCCTCCATCACCGCCATCAAGTTCGCTGGCAACAGAGACATCCAGATG

ATCAGCTGTGGGGCTGACAAGAGCATCTACTTTCGCAGTGCCCAGCAGGGTTCGGATGGACTACACTTTGTCCGTAC

CCACCACGTAGCAGAGAAAACCACCTTGTATGACATGGACATTGACATCACCCAGAAGTACGTGGCCGTGGCCTGCC

AGGACCGCAATGTGAGAGTCTACAACACTGTGAACGGGAAGCAGAAGAAGTGCTACAAGGGCTCCCAGGGTGACGAA

GGGTCCTTGCTGAAGGTCCATGTGGACCCCTCAGGCACCTTCCTGGCCACCAGCTGCTCTGACAAAAGCATCTCAGT

GATTGACTTTTACTCGGGCGAGTGCATTGCCAAGATGTTTGGCCATTCAGAAATTATTACCAGCATGAAGTTCACCT

ATGACTGTCATCACTTGATCACAGTATCTGGAGACAGCTGCGTGTTCATCTGGCACCTGGGCCCGGAGATCACCAAC

TGCATGAAGCAGCACTTGCTGGAGATTGACCACCGGCAGCAGCAGCAGCACACAAATGACAAGAAGCGGAGTGGCCA

CCCCAGGCAGGATACGTATGTGTCCACACCTAGTGAGATTCACTCCCTGAGCCCTGGAGAGCAAACAGAGGATGATC

TGGAGGAAGAGTGTGAGCCAGAAGAGATGCTGAAGACACCATCCAAAGATAGCTTGGATCCAGATCCTCGTTGCCTG

CTAACCAACGGCAAGCTGCCACTGTGGGCAAAGCGGCTGCTAGGGGACGATGATGTGGCAGATGGCTCGGCCTTCCA

CGCCAAGCGCAGCTACCAGCCCCACGGCCGCTGGGCAGAGCGGGCCGGCCAAGAGCCCCTCAAGACCATCCTGGATG
CCCAGGACCTGGATTGCTACTTTACCCCCATGAAGCCCGAGAGTCTGGAGAACTCCATTCTGGATTCACTGGAGCCA
CAGAGCCTGGCCAGCCTGCTGAGTGAGTCAGAGAGTCCCCAGGAAGCTGGCCGCGGGCACCCCTCCTTCCTGCCCCA
GCAGAAGGAATCATCTGAGGCCAGTGAGCTCATCCTCTACTCTCTGGAGGCAGAAGTGACAGTCACAGGGACAGACA
GCCAGTATTGCAGGAAGGAGGTGGAGGCCGGGCCTGGAGACCAGCAGGGCGACTCCTACCTCAGGGTGTCCTCCGAC
AGCCCAAAGGACCAGAGCCCGCCTGAGGACTCGGGGGAGTCAGAGGCCGACCTGGAGTGCAGCTTCGCAGCCATCCA
CTCCCCAGCTCCGCCTCCTGACCCTGCCCCTCGGTTTGCCACGTCGCTGCCCCATTTCCCAGGATGCGCAGGTCCCA
CAGAAGATGAGCTGTCCCTGCCCGAGGGACCCAGCGTCCCCAGCAGCTCCCTACCCCAGACTCCGGAGCAGGAGAAG
TTCCTCCGCCACCACTTTGAGACACTGACTGAGTCCCCCTGCAGAGAGCTCTTCCCCGCAGCTCTGGGAGACGTGGA
GGCCTCTGAAGCTGAAGACCACTTCTTCAACCCACGCCTGAGTATCTCCACGCAGTTCCTCTCAAGCCTCCAGAAGG
CATCCAGGTTCACCCATACCTTCCCTCCCCGGGCAACCCAGTGCCTTGTGAAGTCTCCAGAGGTCAAGCTCATGGAC
CGAGGCGGAAGCCAGCCCAGAGCAGGTACTGGCTACGCCTCCCCAGACAGGACCCACGTCCTCGCTGCAGGGAAGGC
TGAAGAGACCCTGGAGGCCTGGCGCCCACCACCTCCCTGCCTTACGAGCCTGGCGTCCTGTGTCCCTGCTTCCTCCG
TGCTGCCCACAGACAGGAATCTCCCAACGCCCACATCTGCACCCACCCCAGGCCTGGCTCAGGGTGTCCATGCCCCC
TCCACCTGTTCCTACATGGAGGCCACTGCCAGCTCCCGTGCCAGGATATCACGCAGCATCTCCCTCGGTGACAGTGA
GGGCCCTATCGTGGCCACACTGGCCCAGCCCCTCCGTAGGCCATCGTCCGTTGGGGAGCTGGCCTCCTTGGGCCAGG
AGCTTCAGGCCATCACCACCGCGACAACACCCAGTTTGGACAGTGAGGGCCAAGAGCCTGCCCTGCGTTCCTGGGGC
AACCACGAGGCCCGGGCCAACCTGAGACTGACCCTGTCAAGTGCCTGTGATGGGCTCCTGCAGCCCCCCGTGGATAC
CCAGCCTGGCGTCACCGTCCCTGCAGTGAGCTTCCCAGCCCCTAGCCCTGTGGAAGAGAGCGCCCTGAGGCTCCACG
GCTCTGCCTTTCGCCCAAGTCTCCCAGCTCCTGAGTCCCCTGGCCTTCCTGCCCACCCCAGTAACCCCCAGCTTCCA
GAGGCCCGGCCTGGCATCCCTGGCGGCACTGCCTCCCTCCTGGAGCCCACCTCCGGTGCACTTGGTCTGTTCCAGGG
CAGCCCTGCCCGCTGGAGTGAGCCCTGGGTGCCGGTTGAAGCCCTGCCCCCATCTCCCCTTGAGCTGAGCAGGGTGG
GGAACATCTTGCACAGGCTGCAGACCACCTTCCAAGAAGCCCTCGACCTTTACCGTGTGTTGGTCTCCAGTGGCCAG
GTGGACACCGGGCAGCAGCAGGCACGGACTGAGCTGGTCTCCACCTTCCTGTGGATCCACAGCCAGCTGGAGGCTGA
ATGCCTGGTGGGGACTAGTGTGGCCCCAGCCCAGGCTCTGCCCAGCCCAGGACCCCCGTCCCCACCGACGCTGTACC
CCCTGGCCAGCCCAGACCTGCAGGCCCTGCTGGAACACTACTCGGAGCTGCTGGTGCAGGCCGTGCGGAGGAAGGCA
CGGGGGCACTGAGGGCGCAGCCCCTCCACCGCAGCCCTGCTGCTTCTGAGGACTTAGGTATTTTAAGCGAATAAACT
GACAGCTTTGAGGAATGGTTCCTG

SEQ ID NO:1562

>AA281370 # transcript_5 #len 4671 (Includes node 31 - TAA seg 10; node 56)

CGGCGTGACGATGGCGGCCGTAGGGTCCGGAGGCTATGCGCGGAACGATGCAGGGGAGAAGCTGCCCTCTGTCATGG
CGGGAGTTCCGGCGCGGAGGGGCCAGTCCTCCCCGCCCCCCGCCCCACCAATCTGCCTACGGCGGCGGACGCGACTC
TCGACGGCCTCCGAGGAGACGGTGCAGAACCGGGTGTCACTCGAGAAGGTGCTTGGCATCACAGCCCAGAACAGCAG
TGGCCTAACCTGTGACCCCGGCACAGGCCATGTGGCCTACCTGGCAGGCTGTGTGGTGGTGATTTTGGACCCCAAGG
AGAACAAGCAGCAGCACATCTTTAACACCGCCAGGAAGTCTCTCAGTGCTCTGGCCTTCTCCCCTGATGGGAAGTAC
ATAGTGACAGGGGAGAATGGGCATAGGCCTGCTGTGCGCATCTGGGATGTGGAGGAGAAGAATCAGGTGGCGGAGAT
GCTAGGCCACAAGTATGGTGTGGCGTGTGTGGCCTTCTCACCCAATATGAAGCACATCGTGTCCATGGGCTACCAAC

ATGACATGGTGCTCAACGTCTGGGACTGGAAGAAAGACATCGTAGTGGCCTCCAACAAGGTATCTTGTAGAGTCATT
GCCCTCTCCTTCTCAGAGGACAGCAGCTATTTTGTCACTGTTGGGAACCGCCATGTGAGGTTCTGGTTCTTGGAAGT
CTCCACTGAGACAAAGGTGACGAGCACAGTGCCCCTTGTAGGGCGCTCGGGCATCCTGGGCGAGCTGCACAACAACA
TCTTCTGTGGTGTGGCCTGCGGTCGGGGCCGGATGGCGGGCAGTACCTTCTGTGTGTCCTACTCGGGCCTCCTCTGC
CAGTTCAATGAGAAGAGGGTGCTGGAGAAGTGGATCAACCTGAAGGTCTCCCTGTCTTCCTGCCTCTGTGTCAGCCA
GGAGCTCATCTTCTGTGGCTGCACAGATGGGATAGTCCGCATCTTCCAGGCCCATAGCCTGCACTACCTCGCCAACC
TGCCCAAGCCACACTACCTTGGGGTAGACGTGGCACAGGGCCTGGAGCCCAGCTTCCTCTTCCACAGGAAGGCGGAA
GCAGTCTACCCAGATACAGTGGCACTGACCTTCGACCCCATCCACCAGTGGCTGTCCTGCGTGTATAAGGACCACAG
CATCTACATCTGGGATGTCAAGGACATCAACAGAGTGGGCAAGGTGTGGTCAGAGCTCTTCCACAGCTCCTACGTTT
GGAACGTGGAGGTGTATCCTGAGTTTGAAGACCAGAGAGCTTGTTTGCCATCAGGATCCTTTCTGACTTGTTCTTCA
GACAACACCATTCGCTTCTGGAACTTGGACAGCAGCCCTGATTCTCACTGGCAGAAAAACATCTTCAGCAATACCCT
GCTGAAGGTCGTGTACGTGGAGAATGACATCCAGCACCTGCAGGACATGTCACACTTCCCAGACCGGGGGAGCGAGA
ATGGGACACCCATGGACGTGAAAGCCGGGGTGCGGGTCATGCAGGTCAGTCCTGACGGCCAGCATTTGGCTTCAGGC
GACCGAAGTGGAAATCTGAGGATCCACGAGCTGCACTTCATGGACGAGCTGGTCAAGGTGGAGGCCCATGATGCTGA
GGTGCTGTGCCTGGAGTACTCCAAGCCAGAGACGGGGCTGACCTTGCTGGCCTCAGCCAGTCGGGACCGGCTGATCC
ATGTGCTGAACGTGGAGAAGAACTACAACCTGGAGCAGACGCTGGATGACCACTCCTCCTCCATCACCGCCATCAAG
TTCGCTGGCAACAGAGACATCCAGATGATCAGCTGTGGGGCTGACAAGAGCATCTACTTTCGCAGTGCCCAGCAGGG
TTCGGATGGACTACACTTTGTCCGTACCCACCACGTAGCAGAGAAAACCACCTTGTATGACATGGACATTGACATCA
CCCAGAAGTACGTGGCCGTGGCCTGCCAGGACCGCAATGTGAGAGTCTACAACACTGTGAACGGGAAGCAGAAGAAG
TGCTACAAGGGCTCCCAGGGTGACGAAGGGTCCTTGCTGAAGGTCCATGTGGACCCCTCAGGCACCTTCCTGGCCAC
CAGCTGCTCTGACAAAAGCATCTCAGTGATTGACTTTTACTCGGGCGAGTGCATTGCCAAGATGTTTGGCCATTCAG
AAATTATTACCAGCATGAAGTTCACCTATGACTGTCATCACTTGATCACAGTATCTGGAGACAGCTGCGTGTTCATC
TGGCACCTGGGCCCGGAGATCACCAACTGCATGAAGCAGCACTTGCTGGAGATTGACCACCGGCAGCAGCAGCAGCA
CACAAATGACAAGAAGCGGAGTGGCCACCCCAGGCAGGATACGTATGTGTCCACACCTAGTGAGATTCACTCCCTGA
GCCCTGGAGAGCAAACAGAGGATGATCTGGAGGAAGAGTGTGAGCCAGAAGAGATGCTGAAGACACCATCCAAAGAT
AGCTTGGATCCAGATCCTCGTTGCCTGCTAACCAACGGCAAGCTGCCACTGTGGGCAAAGCGGCTGCTAGGGGACGA
TGATGTGGCAGATGGCTCGGCCTTCCACGCCAAGCGCAGCTACCAGCCCCACGGCCGCTGGGCAGAGCGGGCCGGCC
AAGAGCCCCTCAAGACCATCCTGGATGCCCAGGACCTGGATTGCTACTTTACCCCCATGAAGCCCGAGAGTCTGGAG
AACTCCATTCTGGATTCACTGGAGCCACAGAGCCTGGCCAGCCTGCTGAGTGAGTCAGAGAGTCCCCAGGAAGCTGG
CCGCGGGCACCCCTCCTTCCTGCCCCAGCAGAAGGAATCATCTGAGGCCAGTGAGCTCATCCTCTACTCTCTGGAGG
CAGAAGTGACAGTCACAGGGACAGACAGCCAGTATTGCAGGAAGGAGGTGGAGGCCGGGCCTGGAGACCAGCAGGGC
GACTCCTACCTCAGGGTGTCCTCCGACAGCCCAAAGGACCAGAGCCCGCCTGAGGACTCGGGGGAGTCAGAGGCCGA
CCTGGAGTGCAGCTTCGCAGCCATCCACTCCCCAGCTCCGCCTCCTGACCCTGCCCCTCGGTTTGCCACGTCGCTGC
CCCATTTCCCAGGATGCGCAGGTCCCACAGAAGATGAGCTGTCCCTGCCCGAGGGACCCAGCGTCCCCAGCAGCTCC
CTACCCCAGACTCCGGAGCAGGAGAAGTTCCTCCGCCACCACTTTGAGACACTGACTGAGTCCCCCTGCAGAGAGCT
CTTCCCCGCAGCTCTGGGAGACGTGGAGGCCTCTGAAGCTGAAGACCACTTCTTCAACCCACGCCTGAGTATCTCCA
CGCAGTTCCTCTCAAGCCTCCAGAAGGCATCCAGGTTCACCCATACCTTCCCTCCCCGGGCAACCCAGTGCCTTGTG
AAGTCTCCAGAGGTCAAGCTCATGGACCGAGGCGGAAGCCAGCCCAGAGCAGGTACTGGCTACGCCTCCCCAGACAG

GACCCACGTCCTCGCTGCAGGGAAGGCTGAAGAGACCCTGGAGGCCTGGCGCCCACCACCTCCCTGCCTTACGAGCC
TGGCGTCCTGTGTCCCTGCTTCCTCCGTGCTGCCCACAGACAGGAATCTCCCAACGCCCACATCTGCACCCACCCCA
GGCCTGGCTCAGGGTGTCCATGCCCCCTCCACCTGTTCCTACATGGAGGCCACTGCCAGCTCCCGTGCCAGGATATC
ACGCAGCATCTCCCTCGGTGACAGTGAGGGCCCTATCGTGGCCACACTGGCCCAGCCCCTCCGTAGGCCATCGTCCG
TTGGGGAGCTGGCCTCCTTGGGCCAGGAGCTTCAGGCCATCACCACCGCGACAACACCCAGTTTGGACAGTGAGGGC
CAAGAGCCTGCCCTGCGTTCCTGGGGCAACCACGAGGCCCGGGCCAACCTGAGACTGACCCTGTCAAGTGCCTGTGA
TGGGCTCCTGCAGCCCCCCGTGGATACCCAGCCTGGCGTCACCGTCCCTGCAGTGAGCTTCCCAGCCCCTAGCCCTG
TGGAAGAGAGCGCCCTGAGGCTCCACGGCTCTGCCTTTCGCCCAAGTCTCCCAGCTCCTGAGTCCCCTGGCCTTCCT
GCCCACCCCAGTAACCCCCAGCTTCCAGAGGCCCGGCCTGGCATCCCTGGCGGCACTGCCTCCCTCCTGGAGCCCAC
CTCCGGTGCACTTGGTCTGTTCCAGGGCAGCCCTGCCCGCTGGAGTGAGCCCTGGGTGCCGGTTGAAGCCCTGCCCC
CATCTCCCCTTGAGCTGAGCAGGGTGGGGAACATCTTGCACAGGCTGCAGACCACCTTCCAAGAAGCCCTCGACCTT
TACCGTGTGTTGGTCTCCAGTGGCCAGGTGGACACCGGGCAGCAGCAGGCACGGACTGAGCTGGTCTCCACCTTCCT
GTGGATCCACAGCCAGCTGGAGGCTGAATGCCTGGTGGGGACTAGTGTGGCCCCAGCCCAGGCTCTGCCCAGCCCAG
GACCCCCGTCCCCACCGACGCTGTACCCCCTGGCCAGCCCAGACCTGCAGGCCCTGCTGGAACACTACTCGGAGCTG
CTGGTGCAGGCCGTGCGGAGGAAGGCACGGGGGCACTGAGGGCGCAGCCCCTCCACCGCAGCCCTGCTGCTTCTGAG
GACTTAGGTATTTTAAGCGAATAAACTGACAGCTTTGAGGAATGGTTCCTG

SEQ ID NO:1563

> AA281370 p1 # trn_0 #len 1044

MSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAEVLCLEYSKPETGLTL
LASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQGSDGLHFVRTHHVAEK
TTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLATSCSDKSISVIDFYSG
ECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQHTNDKKRSGHPRQDTY
VSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDDDVADGSAFHAKRSYQ
PHGRWAERAGQEPLKTILDAQDLDCYFTPMKPESLENSILDSLEPQSLASLLSESESPQEAGRGHPSFLPQQKESSE
ASELILYSLEAEVTVTGTDSQYCRKEVEAGPGDQQGDSYLRVSSDSPKDQSPPEDSGESEADLECSFAAIHSPAPPP
DPAPRFATSLPHFPGCAGPTEDELSLPEGPSVPSSSLPQTPEQEKFLRHHFETLTESPCRALGDVEASEAEDHFFNP
RLSISTQFLSSLQKASRFTHTFPPRATQCLVKSPEVKLMDRGGSQPRAGTGYASPDRTHVLAAGKAEETLEAWRPPP
PCLTSLASCVPASSVLPTDRNLPTPTSAPTPGLAQGVHAPSTCSYMEATASSRARISRSISLGDSEGPIVATLAQPL
RRPSSVGELASLGQELQAITTATTPSLDSEGQEPALRSWGNHEARANLRLTLSSACDGLLQPPVDTQPGVTVPAVSF
PAPSPVEESALRLHGSAFRPSLPAPESPGLPAHPSNPQLPEARPGIPGGTASLLEPTSGALGLFQGSPARWSEPWVP
VEALPPSPLELSRVGNILHRLQTTFQEALDLYRVLVSSGQVDTGQQQARTELVSTFLWIHSQLEAECLVGTSVAPAQ
ALPSPGPPSPPTLYPLASPDLQALLEHYSELLVQAVRRKARGH

SEQ ID NO:1564

> AA281370 p2 # trn_1 #len 1521

GVTMAAVGSGGYARNDAGEKLPSVMAGVPARRGQSSPPPAPPICLRRRTRLSTASEETVQNRVSLEKVLGITAQNSS
GLTCDPGTGHVAYLAGCVVVILDPKENKQQHIFNTARKSLSALAFSPDGKYIVTGENGHRPAVRIWDVEEKNQVAEM

LGHKYGVACVAFSPNMKHIVSMGYQHDMVLNVWDWKKDIVVASNKVSCRVIALSFSEDSSYFVTVGNRHVRFWFLEV
STETKVTSTVPLVGRSGILGELHNNIFCGVACGRGRMAGSTFCVSYSGLLCQFNEKRVLEKWINLKVSLSSCLCVSQ
ELIFCGCTDGIVRIFQAHSLHYLANLPKPHYLGVDVAQGLEPSFLFHRKAEAVYPDTVALTFDPIHQWLSCVYKDHS
IYIWDVKDINRVGKVWSELFHSSYVWNVEVYPEFEDQRACLPSGSFLTCSSDNTIRFWNLDSSPDSHWQKNIFSNTL
LKVVYVENDIQHLQDMSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAE
VLCLEYSKPETGLTLLASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQG
SDGLHFVRTHHVAEKTTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLAT
SCSDKSISVIDFYSGECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQH
TNDKKRSGHPRQDTYVSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDD
DVADGSAFHAKRSYQPHGRWAERAGQEPLKTILDAQDLDCYFTPMKPESLENSILDSLEPQSLASLLSESESPQEAG
RGHPSFLPQQKESSEASELILYSLEAEVTVTGTDSQYCRKEVEAGPGDQQGDSYLRVSSDSPKDQSPPEDSGESEAD
LECSFAAIHSPAPPPDPAPRFATSLPHFPGCAGPTEDELSLPEGPSVPSSSLPQTPEQEKFLRHHFETLTESPCRAL
GDVEASEAEDHFFNPRLSISTQFLSSLQKASRFTHTFPPRATQCLVKSPEVKLMDRGGSQPRAGTGYASPDRTHVLA
AGKAEETLEAWRPPPPCLTSLASCVPASSVLPTDRNLPTPTSAPTPGLAQGVHAPSTCSYMEATASSRARISRSISL
GDSEGPIVATLAQPLRRPSSVGELASLGQELQAITTATTPSLDSEGQEPALRSWGNHEARANLRLTLSSACDGLLQP
PVDTQPGVTVPAVSFPAPSPVEESALRLHGSAFRPSLPAPESPGLPAHPSNPQLPEARPGIPGGTASLLEPTSGALG
LFQGSPARWSEPWVPVEALPPSPLELSRVGNILHRLQTTFQEALDLYRVLVSSGQVDTGQQQARTELVSTFLWIHSQ
LEAECLVGTSVAPAQALPSPGPPSPPTLYPLASPDLQALLEHYSELLVQAVRRKARGH

SEQ ID NO:1565

> AA281370 p5 # trn_4 #len 1049

MSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAEVLCLEYSKPETGLTL
LASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQGSDGLHFVRTHHVAEK
TTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLATSCSDKSISVIDFYSG
ECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQHTNDKKRSGHPRQDTY
VSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDDDVADGSAFHAKRSYQ
PHGRWAERAGQEPLKTILDAQDLDCYFTPMKPESLENSILDSLEPQSLASLLSESESPQEAGRGHPSFLPQQKESSE
ASELILYSLEAEVTVTGTDSQYCRKEVEAGPGDQQGDSYLRVSSDSPKDQSPPEDSGESEADLECSFAAIHSPAPPP
DPAPRFATSLPHFPGCAGPTEDELSLPEGPSVPSSSLPQTPEQEKFLRHHFETLTESPCRELFPAALGDVEASEAED
HFFNPRLSISTQFLSSLQKASRFTHTFPPRATQCLVKSPEVKLMDRGGSQPRAGTGYASPDRTHVLAAGKAEETLEA
WRPPPPCLTSLASCVPASSVLPTDRNLPTPTSAPTPGLAQGVHAPSTCSYMEATASSRARISRSISLGDSEGPIVAT
LAQPLRRPSSVGELASLGQELQAITTATTPSLDSEGQEPALRSWGNHEARANLRLTLSSACDGLLQPPVDTQPGVTV
PAVSFPAPSPVEESALRLHGSAFRPSLPAPESPGLPAHPSNPQLPEARPGIPGGTASLLEPTSGALGLFQGSPARWS
EPWVPVEALPPSPLELSRVGNILHRLQTTFQEALDLYRVLVSSGQVDTGQQQARTELVSTFLWIHSQLEAECLVGTS
VAPAQALPSPGPPSPPTLYPLASPDLQALLEHYSELLVQAVRRKARGH

SEQ ID NO:1566

> AA281370 p6 # trn_5 #len 1526

GVTMAAVGSGGYARNDAGEKLPSVMAGVPARRGQSSPPPAPPICLRRRTRLSTASEETVQNRVSLEKVLGITAQNSS
GLTCDPGTGHVAYLAGCVVVILDPKENKQQHIFNTARKSLSALAFSPDGKYIVTGENGHRPAVRIWDVEEKNQVAEM
LGHKYGVACVAFSPNMKHIVSMGYQHDMVLNVWDWKKDIVVASNKVSCRVIALSFSEDSSYFVTVGNRHVRFWFLEV
STETKVTSTVPLVGRSGILGELHNNIFCGVACGRGRMAGSTFCVSYSGLLCQFNEKRVLEKWINLKVSLSSCLCVSQ
ELIFCGCTDGIVRIFQAHSLHYLANLPKPHYLGVDVAQGLEPSFLFHRKAEAVYPDTVALTFDPIHQWLSCVYKDHS
IYIWDVKDINRVGKVWSELFHSSYVWNVEVYPEFEDQRACLPSGSFLTCSSDNTIRFWNLDSSPDSHWQKNIFSNTL
LKVVYVENDIQHLQDMSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAE
VLCLEYSKPETGLTLLASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQG
SDGLHFVRTHHVAEKTTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLAT
SCSDKSISVIDFYSGECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQH
TNDKKRSGHPRQDTYVSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDD
DVADGSAFHAKRSYQPHGRWAERAGQEPLKTILDAQDLDCYFTPMKPESLENSILDSLEPQSLASLLSESESPQEAG
RGHPSFLPQQKESSEASELILYSLEAEVTVTGTDSQYCRKEVEAGPGDQQGDSYLRVSSDSPKDQSPPEDSGESEAD
LECSFAAIHSPAPPPDPAPRFATSLPHFPGCAGPTEDELSLPEGPSVPSSSLPQTPEQEKFLRHHFETLTESPCREL
FPAALGDVEASEAEDHFFNPRLSISTQFLSSLQKASRFTHTFPPRATQCLVKSPEVKLMDRGGSQPRAGTGYASPDR
THVLAAGKAEETLEAWRPPPPCLTSLASCVPASSVLPTDRNLPTPTSAPTPGLAQGVHAPSTCSYMEATASSRARIS
RSISLGDSEGPIVATLAQPLRRPSSVGELASLGQELQAITTATTPSLDSEGQEPALRSWGNHEARANLRLTLSSACD
GLLQPPVDTQPGVTVPAVSFPAPSPVEESALRLHGSAFRPSLPAPESPGLPAHPSNPQLPEARPGIPGGTASLLEPT
SGALGLFQGSPARWSEPWVPVEALPPSPLELSRVGNILHRLQTTFQEALDLYRVLVSSGQVDTGQQQARTELVSTFL
WIHSQLEAECLVGTSVAPAQALPSPGPPSPPTLYPLASPDLQALLEHYSELLVQAVRRKARGH

SEQ ID NO:1567
>AA281370 # node_31 (TAA seg 10) #len 122
GGTTCGGATGGACTACACTTTGTCCGTACCCACCACGTAGCAGAGAAAACCACCTTGTAT
GACATGGACATTGACATCACCCAGAAGTACGTGGCCGTGGCCTGCCAGGACCGCAATGTG
AG

SEQ ID NO:1568
>AA281370 # node_56 (no TAA seg) #len 15
AGCTCTTCCCCGCAG

SEQ ID NO:1569
>AA281370 - Unique aa seq coded by node 56
ELFPA

SEQ ID NO:1570

>AA281370 - The elongated unique aa seq not found in SwissProt (Q96AD9)encoded by T1 and 5. Note:even after the elongation the protein is probably partial.

GVTMAAVGSGGYARNDAGEKLPSVMAGVPARRGQSSPPPAPPICLRRRTRLSTASEETVQNRVSLEKVLGITAQNSS
GLTCDPGTGHVAYLAGCVVVILDPKENKQQHIFNTARKSLSALAFSPDGKYIVTGENGHRPAVRIWDVEEKNQVAEM
LGHKYGVACVAFSPNMKHIVSMGYQHDMVLNVWDWKKDIVVASNKVSCRVIALSFSEDSSYFVTVGNRHVRFWFLEV
STETKVTSTVPLVGRSGILGELHNNIFCGVACGRGRMAGSTFCVSYSGLLCQFNEKRVLEKWINLKVSLSSCLCVSQ
ELIFCGCTDGIVRIFQAHSLHYLANLPKPHYLGVDVAQGLEPSFLFHRKAEAVYPDTVALTFDPIHQWLSCVYKDHS
IYIWDVKDINRVGKVWSELFHSSYVWNVEVYPEFEDQRACLPSGSFLTCSSDNTIRFWNLDSSPDSHWQKNIFSNTL
LKVVYVENDIQHLQDMSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAE
VLCLEYSKPETGLTLLASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQG
SDGLHFVRTHHVAEKTTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLAT
SCSDKSISVIDFYSGECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQH
TNDKKRSGHPRQDTYVSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDD
DVADGSAFHAKRSYQPHGRWAERAGQEPLKTILDAQDLDCYFTP

SEQ ID NO:1571

>AA281370 - The elongated unique aa seq not found in SwissProt (Q96AD9)encoded by T0 and 4.

MSHFPDRGSENGTPMDVKAGVRVMQVSPDGQHLASGDRSGNLRIHELHFMDELVKVEAHDAEVLCLEYSKPETGLTL
LASASRDRLIHVLNVEKNYNLEQTLDDHSSSITAIKFAGNRDIQMISCGADKSIYFRSAQQGSDGLHFVRTHHVAEK
TTLYDMDIDITQKYVAVACQDRNVRVYNTVNGKQKKCYKGSQGDEGSLLKVHVDPSGTFLATSCSDKSISVIDFYSG
ECIAKMFGHSEIITSMKFTYDCHHLITVSGDSCVFIWHLGPEITNCMKQHLLEIDHRQQQQHTNDKKRSGHPRQDTY
VSTPSEIHSLSPGEQTEDDLEEECEPEEMLKTPSKDSLDPDPRCLLTNGKLPLWAKRLLGDDDVADGSAFHAKRSYQ
PHGRWAERAGQEPLKTILDAQDLDCYFTP

SEQ ID NO:1572

>Forward primer:

ACTCACTCAGAGACTAACACAAAGGAAG

SEQ ID NO:1573

>Reverse primer:

AGTATGGGAAGAATTTACTGGTCACA

SEQ ID NO:1574

>Amplicon :

ACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTACAATAAGTTCATCCTTCTT
CAGTGTGACCAGTAAATTCTTCCCATACT

SEQ ID NO:1575

>Z21368 # transcript_4 #len 3483 (includes node 51 - TAA-seg 19)

GGCCATTCCAATGGAACAGACAGGGTAAGGACCAATCTGGACTGTGTTATCTTTTCCAGGTGCAAGTATGTGCTATG

GAACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTATGCAGTACACAGGACCAATGCTGCCCATC

CACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCT

GTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCATTTACACCGCCGACCATGGTTACCATATTG

GGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCA

AGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGG

GCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTC

GAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGGCAAATTTCTACGTAAGAAGGAAGAA

TCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCAAAGAACTATGCCAGCAGGCCAGGTA

CCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAATGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGT

GTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGCACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGAC

AAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAA

CCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTG

AAATATATGACATAAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGAT

GAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAA

CGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACAAGTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGA

GAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATAAGGCATACATTGACAAAGAGATTGAAGCTCTGCAAGAT

AAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATGTAGCTGCAGTAAACAAAG

CTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCCATTCAAGGAGGCTGCTC

AGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGAGGAAGGAGAAGAGACGG

CAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAACCACTGGCAGACAGCCCC

GTTCTGGAACCTGGGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATG

AGACGCATAATTTTCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAG

CTCACAAATACAGTGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTG

TCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATCTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTAC

ACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGA

GGAGCTACACAGTGTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATT

ACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACT

GCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGA

AGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGAC

CAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGG

AGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGA

GCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTG

GAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATG

GCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGA

GAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAAC
ACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGG
GTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTA
TCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTG
TACTAAAACAGTATTATCTTTTGAATATCGTAGGGACATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATG
GTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGTAAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTT
TTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAAT
GCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAA
AGTCAAATAAACCCCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTT
GCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATA
AATAAGTAAACAAAATGA

SEQ ID NO:1576

>Z21368 # transcript_5 #len 3344 (includes node 53 - TAA seg 27)

GGCTGCATCTTGCCTGGTGATGTGGTCAGAATACAGGGGTGCAGGCATCTCTCCAGCCTGACCCTGGCAAGAGTCAG
TTAATCTTGCTCAGTGCCATTGCTGTGATCACACACCCACCCTTGCCACACAACTACCCATGCCTAGGAGACCAGAT
GAGAGGGTGAGAAGAGTTGAAGGCCAATGAGTCACTGCTGTAGAAAAAGCAGCCCTAAGTGCCACCTTCCCCTGGCA
TTGGATCTCAGCCATCACCGTGTGCCCCTTTACAGAGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACGAT
CCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAAGC
CAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGGCAAATTTCTA
CGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCAAAGAACTATG
CCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAATGCATTGAGGATACATCTGGCAAGC
TTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGCACGCGGAACCTCTACGCTCGCGGC
TTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAACG
GCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTCCG
TCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACATT
GCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATGCT
GGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACAAGTGTTTTATTCTTCCCAATGACT
CTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATAAGGCATACATTGACAAAGAGATT
GAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATGTAG
CTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCCAT
TCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGAGG
AAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAACCA
CTGGCAGACAGCCCCGTTCTGGAACCTGGGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTGTT
TGCGTACAGTTAATGAGACGCATAATTTTCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGAAT
ACAGATCCTTATCAGCTCACAAATACAGTGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTAAT
GGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATCTTGATGTTGGAAATAAAGATGGAG
GAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATCAA

CTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCTGG
TGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAACA
AATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCCTC
TGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAACCC
TGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATTTG
AATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGTGG
CGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAGAACTTCCCC
AGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATTTCAGTTCAT
CAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAAGGTGGAGAA
GAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGATGAAACTGT
TACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGCCACCAACAT
TCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACGGAGACTCAT
CGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGATTTTTTGCT
TGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGACATAAGTATATACATGTTATCCAATC
AAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGTAAATCTTTCAACACACTTCCACTGC
CTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTGCA
CTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTTGG
CTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGCTT
TTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTTGG
TGGTGAAAATAAATAAATAAGTAAACAAAATGA

SEQ ID NO:1577
>Z21368 # transcript_12 #len 1247 (includes node 52 - TAA seg 21 ; node 25 - TAA seg 32)
AGCAACACATTCAGGCAAAGGGATGCGAGAAGAACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGAT
TATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTT
TGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATC
ATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGA
TATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATGTTTCGAACAAACAAGAAGGCCAAAA
TTTGGCGTGATACATTCCTAGTGGAAAGAGGGTAATTATTGGTTCCTGGGGTGCTTCTGGGAACCAGTCCTAGTGGG
CAGCTTTCCCTGCTGAGTATTTTTTTTCTCCTTATTTTTGTTTACTAAGCATGCAGATTTCGTAAACCTAGTCACAA
GATTGAATGGTTTGCTGCTTATTCTGTAGTGGTCAATAGAGTAATAATTGCTGGATCAGAATTGTAAAGAATAACCC
TCAAGTTGGTTAATTGGTACAAAAACACAGTTAGATAGAAGTTATAGAATTTGATAGTATAGTTGGGACATTATCGT
TAACAATAATTTATGTATATCTTAAAATAGCTAGAAGTGAAGAATTGCAAAGTTCCCAACACAAGGAAAAGATAAAT
GAGATGATGAATATCCCAATTATCTTGATTTGATCATTACACATTGTAGACTGGTATCCATATATCACACGTACCCC
CAAAATATGTATAATTGTGATATATCAATTTTTAAAATACCAAAAAAGCAAGAGAATGACGACTCCACATTCCCCAA
AAAGAATAAATTCTCATAAGCTTGGACCAAAGCCTTTATCATGGGTGTAGATTTACTGTTGCATTTCTCAGTGCTGG
TTTCTAATCAGACCAGTGGATTGAGTTTCTCTACCATCCTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACC

CGGCACCCTTCTTCCTACTTCCTACTTCTTTTCCTTGTGTGCCTTTCCTAGTTTTAAATAGATAAATGTATGCCATT
GTAATTATTTCCATTGTCACTTCTGGGTTTCCCCTTTTGGTTCATTAATACCCATTGCCTTGTTTTTCTCTGTACAT
AAATTAGGAGAGAGA

SEQ ID NO:1578
>Z21368 # transcript_13 #len 5710 (includes node 2 - TAA seg 5)
AGGTTACTTGACTGGGAGTTCTCAGACCTCCAGTTTCAGCCCTGCCCTCAGCCTCCAATCCGTAAGAGACACCCAGC
CCCAGCAATTGGATTGGGCAGCCCGTCTTGACACACCACTGTGCTGAGTGCTTGAGGACGTGTTTCAACAGATGGTT
GGGGTTAGTGTGTGTCATCACATTCGAGTGGGGATTAAGAGAAGGAAGGCTGCCTTGCTGGAGCTGTGTGGTCTTCT
CCAAGTGAGAGTCGCAGGCAATAGAACTACTTTGCTTTTGGAGGAAAAGGAGGAATTCATTTTCAGCAGACACAAGA
AAAGCAGTTTTTTTTTCAGGGATTCTTCACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAA
TTTCTTACCTGGTCATTATTTAGTCTACAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCT
TGAAGAGAGCATAATTGGAATGGAGAGGTGCTGACGGCCACCCACCATCATCTAAAGAAGATAAACTTGGCAAATGA
CATGCAGGTTCTTCAAGGCAGAATAATTGCAGAAAATCTTCAAAGGACCCTATCTGCAGATGTTCTGAATACCTCTG
AGAATAGAGATTGATTATTCAACCAGGATACCTAATTCAAGAACTCCAGAAATCAGGAGACGGAGACATTTTGTCAG
TTTTGCAACATTGGACCAAATACAATGAAGTATTCTTGCTGTGCTCTGGTTTTGGCTGTCCTGGGCACAGAATTGCT
GGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGTTCAGAGGACGGATACAGCAGGAACGAAAAAACATCCGACCCA
ACATTATTCTTGTGCTTACCGATGATCAAGATGTGGAGCTGGGGTCCCTGCAAGTCATGAACAAAACGAGAAAGATT
ATGGAACATGGGGGGGCCACCTTCATCAATGCCTTTGTGACTACACCCATGTGCTGCCCGTCACGGTCCTCCATGCT
CACCGGGAAGTATGTGCACAATCACAATGTCTACACCAACAACGAGAACTGCTCTTCCCCCTCGTGGCAGGCCATGC
ATGAGCCTCGGACTTTTGCTGTATATCTTAACAACACTGGCTACAGAACAGCCTTTTTTGGAAAATACCTCAATGAA
TATAATGGCAGCTACATCCCCCCTGGGTGGCGAGAATGGCTTGGATTAATCAAGAATTCTCGCTTCTATAATTACAC
TGTTTGTCGCAATGGCATCAAAGAAAAGCATGGATTTGATTATGCAAAGGACTACTTCACAGACTTAATCACTAACG
AGAGCATTAATTACTTCAAAATGTCTAAGAGAATGTATCCCCATAGGCCCGTTATGATGGTGATCAGCCACGCTGCG
CCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTCTAAACTGTACCCCAATGCTTCCCAACACATAACTCCTAGTTA
TAACTATGCACCAAATATGGATAAACACTGGATTATGCAGTACACAGGACCAATGCTGCCCATCCACATGGAATTTA
CAAACATTCTACAGCGCAAAAGGCTCCAGACTTTGATGTCAGTGGATGATTCTGTGGAGAGGCTGTATAACATGCTC
GTGGAGACGGGGGAGCTGGAGAATACTTACATCATTTACACCGCCGACCATGGTTACCATATTGGGCAGTTTGGACT
GGTCAAGGGGAAATCCATGCCATATGACTTTGATATTCGTGTGCCTTTTTTTATTCGTGGTCCAAGTGTAGAACCAG
GATCAATAGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACGATCCTGGATATTGCTGGGCTCGACACACCT
CCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAAGCCAGGTAACAGGTTTCGAACAAACAAGAA
GGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGGCAAATTTCTACGTAAGAAGGAAGAATCCAGCAAGAATA
TCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCAAAGAACTATGCCAGCAGGCCAGGTACCAGACAGCCTGT
GAACAACCGGGGCAGAAGTGGCAATGCATTGAGGATACATCTGGCAAGCTTCGAATTCACAAGTGTAAAGGACCCAG
TGACCTGCTCACAGTCCGGCAGAGCACGCGGAACCTCTACGCTCGCGGCTTCCATGACAAAGACAAAGAGTGCAGTT
GTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAACGGCAATTCTTGAGAAACCAGGGGACTCCA
AAGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTCCGTCGAATTTGAAGGTGAAATATATGACAT
AAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACATTGCTAAGCGTCATGATGAAGGCCACAAGG

GGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATGCTGGCAGATAGCAGCAACGCCGTGGGCCCA
CCTACCACTGTCCGAGTGACACACAAGTGTTTTATTCTTCCCAATGACTCTATCCATTGTGAGAGAGAACTGTACCA
ATCGGCCAGAGCGTGGAAGGACCATAAGGCATACATTGACAAAGAGATTGAAGCTCTGCAAGATAAAATTAAGAATT
TAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATGTAGCTGCAGTAAACAAAGCTATTACAATAAA
GAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCCATTCAAGGAGGCTGCTCAGGAAGTAGATAG
CAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGAGGAAGGAGAAGAGACGGCAGAGGAAGGGGG
AAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAACCACTGGCAGACAGCCCCGTTCTGGAACCTG
GGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTGTTTGCGTACAGTTAATGAGACGCATAATTT
TCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGAATACAGATCCTTATCAGCTCACAAATACAG
TGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTAATGGAGCTCAGAAGCTGTCAAGGATATAAG
CAGTGCAACCCAAGACCTAAGAATCTTGATGTTGGAAATAAAGATGGAGGAAGCTATGACCTACACAGAGGACAGTT
ATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATCAACTGGCAAGGCCTAGAGGAGCTACACAGT
GTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCTGGTGGACTGGACTAATTACTTGAAGGATTT
AGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAACAAATAAGACTCAAACTGCTCAAAGTGACG
GGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCCTCTGCTGACTCAGATGAAGACCCAAGGCAT
AAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAACCCTGCATTTGAACCGACCAACATTAAGTCC
AGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATTTGAATTCTGAACACTGGAGAAAAACCGAAA
AATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGTGGCGATGGCATGACAGAGCTAGAGCTCGGG
CCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAGAACTTCCCCAGTATGGTGGTCCTGGAAAGGACATTTT
TGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATTTCAGTTCATCAGATGTTCACCATGGCCACCGCAGAAC
ACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAAGGTGGAGAAGAATCACGAAAAGGAGAAGTCACAGCAC
CTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGATGAAACTGTTACCTTACCCTAAACACAGTATTTCTTT
TTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGCCACCAACATTCCAAGCTACCCTGGGTACCTTTGTGCA
GTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACGGAGACTCATCGTTATAATTTACTATCTGCCAAGAGTA
GAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGATTTTTTGCTTGTTTGTTTGTTTTGTACTAAAACAGTA
TTATCTTTTGAATATCGTAGGGACATAAGTATATACATGTTATCCAATCAAGATGGCTAGAATGGTGCCTTTCTGAG
TGTCTAAAACTTGACACCCCTGGTAAATCTTTCAACACACTTCCACTGCCTGCGTAATGAAGTTTTGATTCATTTTT
AACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTGCACTTTGAGATTAAAATGCCATGTCTATTT
GATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTTGGCTGTCATTGTGACAAAGTCAAATAAACC
CCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGCTTTTGCCAGAAAATGTTGCATGTGTTTTAC
CTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTTGGTGGTGAAAATAAATAAATAAGTAAACAA
AATGAAGATTGCCTGCTCTCTCTGTGCCTAGCCTCAAAGCGTTCATCATACATCATACCTTTAAGATTGCTATATTT
TGGGTTATTTTCTTGACAGGAGAAAAAGATCTAAAGATCTTTTATTTTCATCTTTTTTGGTTTTCTTGGCATGACTA
AGAAGCTTAAATGTTGATAAAATATGACTAGTTTTGAATTTACACCAAGAACTTCTCAATAAAAGAAAATCATGAAT
GCTCCACAATTTCAACATACCACAAGAGAAGTTAATTTCTTAACATTGTGTTCTATGATTATTTGTAAGACCTTCAC
CAAGTTCTGATATCTTTTAAAGACATAGTTCAAAATTGCTTTTGAAAATCTGTATTCTTGAAAATATCCTTGTTGTG
TATTAGGTTTTTAAATACCAGCTAAAGGATTACCTCACTGAGTCATCAGTACCCTCCTATTCAGCTCCCCAAGATGA
TGTGTTTTTGCTTACCCTAAGAGAGGTTTTCTTCTTATTTTTAGATAATTCAAGTGCTTAGATAAATTATGTTTTCT

TTAAGTGTTTATGGTAAACTCTTTTAAAGAAAATTTAATATGTTATAGCTGAATCTTTTTGGTAACTTTAAATCTTT
ATCATAGACTCTGTACATATGTTCAAATTAGCTGCTTGCCTGATGTGTGTATCATCGGTGGGATGACAGAACAAACA
TATTTATGATCATGAATAATGTGCTTTGTAAAAAGATTTCAAGTTATTAGGAAGCATACTCTGTTTTTTAATCATGT
ATAATATTCCATGATACTTTTATAGAACAATTCTGGCTTCAGGAAAGTCTAGAAGCAATATTTCTTCAAATAAAAGG
TGTTTAAACTTT

SEQ ID NO:1579
>Z21368 # transcript_14 #len 5663 (includes node 2 - TAA-seg 5)
CGCAGACCGTCGCTAATGAATCTTGGGGCCGGTGTCGGGCCGGGGCGGCTTGATCGGCAACTAGGAAACCCCAGGCG
CAGAGGCCAGGAGCGAGGGCAGCGAGGATCAGAGGCCAGGCCTTCCCGGCTGCCGGCGCTCCTCGGAGGTCAGGGCA
GATGAGGAACATGACTCTCCCCCTTCGGAGGAGGAAGGAAGTCCCGCTGCCACCTTATCTCTGCTCCTCTGCCTCCT
CCCTGTTCCCAGAGCTTTTTCTCTAGAGAAGATTTTGAAGGCGGCTTTTGGATTCTTCACTTCTCTTGAACAAGGAA
CTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTATTTAGTCTACAATAAGTTCATCCTTCTTC
AGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGAATGGAGAGGTGCTGACGGCCACCCACCA
TCATCTAAAGAAGATAAACTTGGCAAATGACATGCAGGTTCTTCAAGGCAGAATAATTGCAGAAAATCTTCAAAGGA
CCCTATCTGCAGATGTTCTGAATACCTCTGAGAATAGAGATTGATTATTCAACCAGGATACCTAATTCAAGAACTCC
AGAAATCAGGAGACGGAGACATTTTGTCAGTTTTGCAACATTGGACCAAATACAATGAAGTATTCTTGCTGTGCTCT
GGTTTTGGCTGTCCTGGGCACAGAATTGCTGGGAAGCCTCTGTTCGACTGTCAGATCCCCGAGGTTCAGAGGACGGA
TACAGCAGGAACGAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATCAAGATGTGGAGCTGGGGTCC
CTGCAAGTCATGAACAAAACGAGAAAGATTATGGAACATGGGGGGGCCACCTTCATCAATGCCTTTGTGACTACACC
CATGTGCTGCCCGTCACGGTCCTCCATGCTCACCGGGAAGTATGTGCACAATCACAATGTCTACACCAACAACGAGA
ACTGCTCTTCCCCCTCGTGGCAGGCCATGCATGAGCCTCGGACTTTTGCTGTATATCTTAACAACACTGGCTACAGA
ACAGCCTTTTTTGGAAAATACCTCAATGAATATAATGGCAGCTACATCCCCCCTGGGTGGCGAGAATGGCTTGGATT
AATCAAGAATTCTCGCTTCTATAATTACACTGTTTGTCGCAATGGCATCAAAGAAAAGCATGGATTTGATTATGCAA
AGGACTACTTCACAGACTTAATCACTAACGAGAGCATTAATTACTTCAAAATGTCTAAGAGAATGTATCCCCATAGG
CCCGTTATGATGGTGATCAGCCACGCTGCGCCCCACGGCCCCGAGGACTCAGCCCCACAGTTTTCTAAACTGTACCC
CAATGCTTCCCAACACATAACTCCTAGTTATAACTATGCACCAAATATGGATAAACACTGGATTATGCAGTACACAG
GACCAATGCTGCCCATCCACATGGAATTTACAAACATTCTACAGCGCAAAAGGCTCCAGACTTTGATGTCAGTGGAT
GATTCTGTGGAGAGGCTGTATAACATGCTCGTGGAGACGGGGGAGCTGGAGAATACTTACATCATTTACACCGCCGA
CCATGGTTACCATATTGGGCAGTTTGGACTGGTCAAGGGGAAATCCATGCCATATGACTTTGATATTCGTGTGCCTT
TTTTTATTCGTGGTCCAAGTGTAGAACCAGGATCAATAGTCCCACAGATCGTTCTCAACATTGACTTGGCCCCCACG
ATCCTGGATATTGCTGGGCTCGACACACCTCCTGATGTGGACGGCAAGTCTGTCCTCAAACTTCTGGACCCAGAAAA
GCCAGGTAACAGGTTTCGAACAAACAAGAAGGCCAAAATTTGGCGTGATACATTCCTAGTGGAAAGAGGCAAATTTC
TACGTAAGAAGGAAGAATCCAGCAAGAATATCCAACAGTCAAATCACTTGCCCAAATATGAACGGGTCAAAGAACTA
TGCCAGCAGGCCAGGTACCAGACAGCCTGTGAACAACCGGGGCAGAAGTGGCAATGCATTGAGGATACATCTGGCAA
GCTTCGAATTCACAAGTGTAAAGGACCCAGTGACCTGCTCACAGTCCGGCAGAGCACGCGGAACCTCTACGCTCGCG
GCTTCCATGACAAAGACAAAGAGTGCAGTTGTAGGGAGTCTGGTTACCGTGCCAGCAGAAGCCAAAGAAAGAGTCAA
CGGCAATTCTTGAGAAACCAGGGGACTCCAAAGTACAAGCCCAGATTTGTCCATACTCGGCAGACACGTTCCTTGTC

CGTCGAATTTGAAGGTGAAATATATGACATAAATCTGGAAGAAGAAGAAGAATTGCAAGTGTTGCAACCAAGAAACA
TTGCTAAGCGTCATGATGAAGGCCACAAGGGGCCAAGAGATCTCCAGGCTTCCAGTGGTGGCAACAGGGGCAGGATG
CTGGCAGATAGCAGCAACGCCGTGGGCCCACCTACCACTGTCCGAGTGACACACAAGTGTTTTATTCTTCCCAATGA
CTCTATCCATTGTGAGAGAGAACTGTACCAATCGGCCAGAGCGTGGAAGGACCATAAGGCATACATTGACAAAGAGA
TTGAAGCTCTGCAAGATAAAATTAAGAATTTAAGAGAAGTGAGAGGACATCTGAAGAGAAGGAAGCCTGAGGAATGT
AGCTGCAGTAAACAAAGCTATTACAATAAAGAGAAAGGTGTAAAAAAGCAAGAGAAATTAAAGAGCCATCTTCACCC
ATTCAAGGAGGCTGCTCAGGAAGTAGATAGCAAACTGCAACTTTTCAAGGAGAACAACCGTAGGAGGAAGAAGGAGA
GGAAGGAGAAGAGACGGCAGAGGAAGGGGGAAGAGTGCAGCCTGCCTGGCCTCACTTGCTTCACGCATGACAACAAC
CACTGGCAGACAGCCCCGTTCTGGAACCTGGGATCTTTCTGTGCTTGCACGAGTTCTAACAATAACACCTACTGGTG
TTTGCGTACAGTTAATGAGACGCATAATTTTCTTTTCTGTGAGTTTGCTACTGGCTTTTTGGAGTATTTTGATATGA
ATACAGATCCTTATCAGCTCACAAATACAGTGCACACGGTAGAACGAGGCATTTTGAATCAGCTACACGTACAACTA
ATGGAGCTCAGAAGCTGTCAAGGATATAAGCAGTGCAACCCAAGACCTAAGAATCTTGATGTTGGAAATAAAGATGG
AGGAAGCTATGACCTACACAGAGGACAGTTATGGGATGGATGGGAAGGTTAATCAGCCCCGTCTCACTGCAGACATC
AACTGGCAAGGCCTAGAGGAGCTACACAGTGTGAATGAAAACATCTATGAGTACAGACAAAACTACAGACTTAGTCT
GGTGGACTGGACTAATTACTTGAAGGATTTAGATAGAGTATTTGCACTGCTGAAGAGTCACTATGAGCAAAATAAAA
CAAATAAGACTCAAACTGCTCAAAGTGACGGGTTCTTGGTTGTCTCTGCTGAGCACGCTGTGTCAATGGAGATGGCC
TCTGCTGACTCAGATGAAGACCCAAGGCATAAGGTTGGGAAAACACCTCATTTGACCTTGCCAGCTGACCTTCAAAC
CCTGCATTTGAACCGACCAACATTAAGTCCAGAGAGTAAACTTGAATGGAATAACGACATTCCAGAAGTTAATCATT
TGAATTCTGAACACTGGAGAAAAACCGAAAAATGGACGGGGCATGAAGAGACTAATCATCTGGAAACCGATTTCAGT
GGCGATGGCATGACAGAGCTAGAGCTCGGGCCCAGCCCCAGGCTGCAGCCCATTCGCAGGCACCCGAAAGAACTTCC
CCAGTATGGTGGTCCTGGAAAGGACATTTTTGAAGATCAACTATATCTTCCTGTGCATTCCGATGGAATTTCAGTTC
ATCAGATGTTCACCATGGCCACCGCAGAACACCGAAGTAATTCCAGCATAGCGGGGAAGATGTTGACCAAGGTGGAG
AAGAATCACGAAAAGGAGAAGTCACAGCACCTAGAAGGCAGCGCCTCCTCTTCACTCTCCTCTGATTAGATGAAACT
GTTACCTTACCCTAAACACAGTATTTCTTTTTAACTTTTTTATTTGTAAACTAATAAAGGTAATCACAGCCACCAAC
ATTCCAAGCTACCCTGGGTACCTTTGTGCAGTAGAAGCTAGTGAGCATGTGAGCAAGCGGTGTGCACACGGAGACTC
ATCGTTATAATTTACTATCTGCCAAGAGTAGAAAGAAAGGCTGGGGATATTTGGGTTGGCTTGGTTTTGATTTTTTG
CTTGTTTGTTTGTTTTGTACTAAAACAGTATTATCTTTTGAATATCGTAGGGACATAAGTATATACATGTTATCCAA
TCAAGATGGCTAGAATGGTGCCTTTCTGAGTGTCTAAAACTTGACACCCCTGGTAAATCTTTCAACACACTTCCACT
GCCTGCGTAATGAAGTTTTGATTCATTTTTAACCACTGGAATTTTTCAATGCCGTCATTTTCAGTTAGATGATTTTG
CACTTTGAGATTAAAATGCCATGTCTATTTGATTAGTCTTATTTTTTTATTTTTACAGGCTTATCAGTCTCACTGTT
GGCTGTCATTGTGACAAAGTCAAATAAACCCCCAAGGACGACACACAGTATGGATCACATATTGTTTGACATTAAGC
TTTTGCCAGAAAATGTTGCATGTGTTTTACCTCGACTTGCTAAAATCGATTAGCAGAAAGGCATGGCTAATAATGTT
GGTGGTGAAAATAAATAAATAAGTAAACAAAATGAAGATTGCCTGCTCTCTCTGTGCCTAGCCTCAAAGCGTTCATC
ATACATCATACCTTTAAGATTGCTATATTTTGGGTTATTTTCTTGACAGGAGAAAAAGATCTAAAGATCTTTTATTT
TCATCTTTTTTGGTTTTCTTGGCATGACTAAGAAGCTTAAATGTTGATAAAATATGACTAGTTTTGAATTTACACCA
AGAACTTCTCAATAAAAGAAAATCATGAATGCTCCACAATTTCAACATACCACAAGAGAAGTTAATTTCTTAACATT
GTGTTCTATGATTATTTGTAAGACCTTCACCAAGTTCTGATATCTTTTAAAGACATAGTTCAAAATTGCTTTTGAAA
ATCTGTATTCTTGAAAATATCCTTGTTGTGTATTAGGTTTTTAAATACCAGCTAAAGGATTACCTCACTGAGTCATC

AGTACCCTCCTATTCAGCTCCCCAAGATGATGTGTTTTTGCTTACCCTAAGAGAGGTTTTCTTCTTATTTTTAGATA
ATTCAAGTGCTTAGATAAATTATGTTTTCTTTAAGTGTTTATGGTAAACTCTTTTAAAGAAAATTTAATATGTTATA
GCTGAATCTTTTTGGTAACTTTAAATCTTTATCATAGACTCTGTACATATGTTCAAATTAGCTGCTTGCCTGATGTG
TGTATCATCGGTGGGATGACAGAACAAACATATTTATGATCATGAATAATGTGCTTTGTAAAAAGATTTCAAGTTAT
TAGGAAGCATACTCTGTTTTTTAATCATGTATAATATTCCATGATACTTTTATAGAACAATTCTGGCTTCAGGAAAG
TCTAGAAGCAATATTTCTTCAAATAAAAGGTGTTTAAACTTT

SEQ ID NO:1580
>Z21368_PEA_2_P2 # trn_4 #len 630
MCYGTPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYIIYTADHG
YHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPG
NRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLR
IHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVE
FEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTHKCFILPNDSI
HCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFK
EAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLR
TVNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDVGNKDGGS
YDLHRGQLWDGWEG

SEQ ID NO:1581
>Unique aa seq coded by Z21368_PEA_2_P2
MCYG

SEQ ID NO:1582
>Z21368_PEA_2_P3 # trn_5 #len 546
MSHCCRKSSPKCHLPLALDLSHHRVPLYRVPQIVLNIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNK
KAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGP
SDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHTRQTRSLSVEFEGEIYD
INLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTHKCFILPNDSIHCERELY
QSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVD
SKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHN
FLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQ
LWDGWEG

SEQ ID NO:1583
>Unique aa seq coded by Z21368_PEA_2_P3
MSHCCRKSSPKCHLPLALDLSHHRVPLYR

SEQ ID NO:1584
>Z21368_PEA_2_P9 # trn_12 #len 139
SNTFRQRDARRTPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELENTYI
IYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSMFRTNKKAKIWRDTFLVERG

SEQ ID NO:1585
>Unique aa seq coded by Z21368_PEA_2_P9
SNTFRQRDARR

SEQ ID NO:1586
>Z21368_PEA_2_P1 # trn_13; trn_14 #len 871
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSLQVMNKTRKIMEHGGATF
INAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAMHEPRTFAVYLNNTGYRTAFFGKYLNEYNGSYIPP
GWREWLGLIKNSRFYNYTVCRNGIKEKHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSA
PQFSKLYPNASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNMLVETGELEN
TYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVLNIDLAPTILDIAGLDTPPDVDGKSV
LKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLRKKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQ
CIEDTSGKLRIHKCKGPSDLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVH
TRQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLADSSNAVGPPTTVRVTH
KCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNLREVRGHLKRRKPEECSCSKQSYYNKEKGVKKQE
KLKSHLHPFKEAAQEVDSKLQLFKENNRRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTS
SNNNTYWCLRTVNETHNFLFCEFATGFLEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQCNPRPKN
LDVGNKDGGSYDLHRGQLWDGWEG

SEQ ID NO:1587
>Z21368 # node 2 - TAA seg 5 #len 162
GGATTCTTCACTTCTCTTGAACAAGGAACTCACTCAGAGACTAACACAAAGGAAGTAATTTCTTACCTGGTCATTAT
TTAGTCTACAATAAGTTCATCCTTCTTCAGTGTGACCAGTAAATTCTTCCCATACTCTTGAAGAGAGCATAATTGGA
ATGGAGAG

SEQ ID NO:1588
>Z21368 # node 51 - TAA seg 19 #len 78
GGCCATTCCAATGGAACAGACAGGGTAAGGACCAATCTGGACTGTGTTATCTTTTCCAGGTGCAAGTATGTGCTATG
G

SEQ ID NO:1589
>Z21368 # node 52 - TAA seg 21 #len 32
AGCAACACATTCAGGCAAAGGGATGCGAGAAG

SEQ ID NO:1590
>Z21368 # node 53 - TAA seg 27 #len 266
GGCTGCATCTTGCCTGGTGATGTGGTCAGAATACAGGGGTGCAGGCATCTCTCCAGCCTGACCCTGGCAAGAGTCAG
TTAATCTTGCTCAGTGCCATTGCTGTGATCACACACCCACCCTTGCCACACAACTACCCATGCCTAGGAGACCAGAT
GAGAGGGTGAGAAGAGTTGAAGGCCAATGAGTCACTGCTGTAGAAAAAGCAGCCCTAAGTGCCACCTTCCCCTGGCA
TTGGATCTCAGCCATCACCGTGTGCCCCTTTACAG

SEQ ID NO:1591
>Z21368 # node 25 - TAA seg 32 #len 831
GTAATTATTGGTTCCTGGGGTGCTTCTGGGAACCAGTCCTAGTGGGCAGCTTTCCCTGCTGAGTATTTTTTTTCTCC
TTATTTTTGTTTACTAAGCATGCAGATTTCGTAAACCTAGTCACAAGATTGAATGGTTTGCTGCTTATTCTGTAGTG
GTCAATAGAGTAATAATTGCTGGATCAGAATTGTAAAGAATAACCCTCAAGTTGGTTAATTGGTACAAAAACACAGT
TAGATAGAAGTTATAGAATTTGATAGTATAGTTGGGACATTATCGTTAACAATAATTTATGTATATCTTAAAATAGC
TAGAAGTGAAGAATTGCAAAGTTCCCAACACAAGGAAAAGATAAATGAGATGATGAATATCCCAATTATCTTGATTT
GATCATTACACATTGTAGACTGGTATCCATATATCACACGTACCCCCAAAATATGTATAATTGTGATATATCAATTT
TTAAAATACCAAAAAAGCAAGAGAATGACGACTCCACATTCCCCAAAAAGAATAAATTCTCATAAGCTTGGACCAAA
GCCTTTATCATGGGTGTAGATTTACTGTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTC
TACCATCCTCCCCACGTTCTTCTCTAAGCTGCCTCCAAGCCTCACCCGGCACCCTTCTTCCTACTTCCTACTTCTTT
TCCTTGTGTGCCTTTCCTAGTTTTAAATAGATAAATGTATGCCATTGTAATTATTTCCATTGTCACTTCTGGGTTTC
CCCTTTTGGTTCATTAATACCCATTGCCTTGTTTTTCTCTGTACATAAATTAGGAGAGAGA

SEQ ID NO:1592
>Forward primer:
GGCGGCGGCAGGAT

SEQ ID NO:1593
>Reverse primer:
GTCGGGAGCGCAGGG

SEQ ID NO:1594
>Amplicon:
GGCGGCGGCAGGATCGGCCAGAGGAGGAGGGAAGCGCTTTTTTTGATCCTGATTCCAGTTTGCCTCTCTCTTTTTTT
CCCCCAAATTATTCTTCGCCTGATTTTCCTCGCGGAGCCCTGCGCTCCCGAC

SEQ ID NO:1595
>HUMHMGBOX # transcript_0 #len 2533 (Includes node 0 - TAA seg 2)

CCCGTCACATGCGATGGTTGTCTATTAACTTGTTCAAAAAAGTATCAGGAGTTGTCAAGGCAGAGAAGAGAGTGTTT
GCAAAAGGGGGAAAGTAGTTTGCTGCCTCTTTAAGACTAGGACTGAGAGAAAGAAGAGGAGAGAGAAAGAAAGGGAG
AGAAGTTTGAGCCCCAGGCTTAAGCCTTTCCAAAAAATAATAATAACAATCATCGGCGGCGGCAGGATCGGCCAGAG
GAGGAGGGAAGCGCTTTTTTTGATCCTGATTCCAGTTTGCCTCTCTCTTTTTTTCCCCCAAATTATTCTTCGCCTGA
TTTTCCTCGCGGAGCCCTGCGCTCCCGACACCCCCGCCCGCCTCCCCTCCTCCTCTCCCCCCGCCCGCGGGCCCCCC
AAAGTCCCGGCCGGGCCGAGGGTCGGCGGCCGCCGGCGGGCCGGGCCCGCGCACAGCGCCCGCATGTACAACATGAT
GGAGACGGAGCTGAAGCCGCCGGGCCCGCAGCAAACTTCGGGGGGCGGCGGCGGCAACTCCACCGCGGCGGCGGCCG
GCGGCAACCAGAAAAACAGCCCGGACCGCGTCAAGCGGCCCATGAATGCCTTCATGGTGTGGTCCCGCGGGCAGCGG
CGCAAGATGGCCCAGGAGAACCCCAAGATGCACAACTCGGAGATCAGCAAGCGCCTGGGCGCCGAGTGGAAACTTTT
GTCGGAGACGGAGAAGCGGCCGTTCATCGACGAGGCTAAGCGGCTGCGAGCGCTGCACATGAAGGAGCACCCGGATT
ATAAATACCGGCCCCGGCGGAAAACCAAGACGCTCATGAAGAAGGATAAGTACACGCTGCCCGGCGGGCTGCTGGCC
CCCGGCGGCAATAGCATGGCGAGCGGGGTCGGGGTGGGCGCCGGCCTGGGCGCGGGCGTGAACCAGCGCATGGACAG
TTACGCGCACATGAACGGCTGGAGCAACGGCAGCTACAGCATGATGCAGGACCAGCTGGGCTACCCGCAGCACCCGG
GCCTCAATGCGCACGGCGCAGCGCAGATGCAGCCCATGCACCGCTACGACGTGAGCGCCCTGCAGTACAACTCCATG
ACCAGCTCGCAGACCTACATGAACGGCTCGCCCACCTACAGCATGTCCTACTCGCAGCAGGGCACCCCTGGCATGGC
TCTTGGCTCCATGGGTTCGGTGGTCAAGTCCGAGGCCAGCTCCAGCCCCCCTGTGGTTACCTCTTCCTCCCACTCCA
GGGCGCCCTGCCAGGCCGGGGACCTCCGGGACATGATCAGCATGTATCTCCCCGGCGCCGAGGTGCCGGAACCCGCC
GCCCCCAGCAGACTTCACATGTCCCAGCACTACCAGAGCGGCCCGGTGCCCGGCACGGCCATTAACGGCACACTGCC
CCTCTCACACATGTGAGGGCCGGACAGCGAACTGGAGGGGGGAGAAATTTTCAAAGAAAAACGAGGGAAATGGGAGG
GGTGCAAAAGAGGAGAGTAAGAAACAGCATGGAGAAAACCCGGTACGCTCAAAAAGAAAAAGGAAAAAAAAAAATCC
CATCACCCACAGCAAATGACAGCTGCAAAAGAGAACACCAATCCCATCCACACTCACGCAAAAACCGCGATGCCGAC
AAGAAAACTTTTATGAGAGAGATCCTGGACTTCTTTTTGGGGGACTATTTTTGTACAGAGAAAACCTGGGGAGGGTG
GGGAGGGCGGGGGAATGGACCTTGTATAGATCTGGAGGAAAGAAAGCTACGAAAAACTTTTTAAAAGTTCTAGTGGT
ACGGTAGGAGCTTTGCAGGAAGTTTGCAAAAGTCTTTACCAATAATATTTAGAGCTAGTCTCCAAGCGACGAAAAAA
ATGTTTTAATATTTGCAAGCAACTTTTGTACAGTATTTATCGAGATAAACATGGCAATCAAAATGTCCATTGTTTAT
AAGCTGAGAATTTGCCAATATTTTTCAAGGAGAGGCTTCTTGCTGAATTTTGATTCTGCAGCTGAAATTTAGGACAG
TTGCAAACGTGAAAAGAAGAAAATTATTCAAATTTGGACATTTTAATTGTTTAAAAATTGTACAAAAGGAAAAAATT
AGAATAAGTACTGGCGAACCATCTCTGTGGTCTTGTTTAAAAAGGGCAAAAGTTTTAGACTGTACTAAATTTTATAA
CTTACTGTTAAAAGCAAAAATGGCCATGCAGGTTGACACCGTTGGTAATTTATAATAGCTTTTGTTCGATCCCAACT
TTCCATTTTGTTCAGATAAAAAAAACCATGAAATTACTGTGTTTGAAATATTTTCTTATGGTTTGTAATATTTCTGT
AAATTTATTGTGATATTTTAAGGTTTTCCCCCCTTTATTTTCCGTAGTTGTATTTTAAAAGATTCGGCTCTGTATTA
TTTGAATCAGTCTGCCGAGAATCCATGTATATATTTGAACTAATATCATCCTTATAACAGGTACATTTTCAACTTAA
GTTTTTACTCCATTATGCACAGTTTGAGATAAATAAATTTTTGAAATATGGACACTGAAAAAATCTTGA

SEQ ID NO:1596

>HUMHMGBOX_P1 # trn_0 #len 317

MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSEISKRLGA
EWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVN

QRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQG
TPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTAI
NGTLPLSHM

SEQ ID NO:1597
>HUMHMGBOX # node_0 (TAA segment 2) #len 161
TCGGCGGCGGCAGGATCGGCCAGAGGAGGAGGGAAGCGCTTTTTTTGATCCTGATTCCAGTTTGCCTCTCTCTTTTT
TTCCCCCAAATTATTCTTCGCCTGATTTTCCTCGCGGAGCCCTGCGCTCCCGACACCCCCGCCCGCCTCCCCTCCTC
CTCTCCC

SEQ ID NO:1598
>Forward primer:
CCCCAGACTCTGTGCACTTCA

SEQ ID NO:1599
>Reverse primer:
TGGGCTCTGCTCTGTCTTAGTGTA

SEQ ID NO:1600
>Amplicon:
CCCCAGACTCTGTGCACTTCAGACCAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTATGAGAAAACCTGTGTGGACAT
CCCTTGGTGTACACTAAGACAGAGCAGAGCCCA

SEQ ID NO:1601
>HSB6PR # transcript_0 #len 5390 (Includes node 29 - TAA seg 34)
GGGGTGGTGCAGGGCAGGGGTGGTATATCCTGTCTGACGGAGGGCGGGCCTCGCCAGTGCCAGAGAGGGACGAACCA
GGGTGGAAGCGCCAGGAGCAGCTGCAGGGAGCCCTCACGCGGACCTCGCACTCTATGGCCGTAGGGAGCCGCTGAGA
GCGAGAAGAGCACGCTCCTGCCCGCCCGCTGCACCGCACCTCGCCTCGCCTCTCTGCTCTCCTAGGCCCCGGCCGCG
CGCCACCCGCCTCCCGCCACCATGAACCACTCGCCGCTCAAGACCGCCTTGGCGTACGAATGCTTCCAGGACCAGGA
CAACTCCACGTTGGCTTTGCCGTCGGACCAAAAGATGAAAACAGGCACGTCTGGCAGGCAGCGCGTGCAGGAGCAGG
TGATGATGACCGTCAAGCGGCAGAAGTCCAAGTCTTCCCAGTCGTCCACCCTGAGCCACTCCAATCGAGGTTCCATG
TATGATGGCTTGGCTGACAATTACAACTATGGGACCACCAGCAGGAGCAGCTACTACTCCAAGTTCCAGGCAGGGAA
TGGCTCATGGGGATATCCGATCTACAATGGAACCCTCAAGCGGGAGCCTGACAACAGGCGCTTCAGCTCCTACAGCC
AGATGGAGAACTGGAGCCGGCACTACCCCCGGGGGCAGCTGTAACACCACCGGCGCAGGCAGCGACATCTGCTTCATG
CAGAAAATCAAGGCGAGCCGCAGTGAGCCCGACCTCTACTGTGACCCACGGGGCACCCTGCGCAAGGGCACGCTGGG
CAGCAAGGGCCAGAAGACCACCCAGAACCGCTACAGCTTTTACAGCACCTGCAGTGGTCAGAAGGCCATAAAGAAGT
GCCCTGTGCGCCCGCCCTCTTGTGCCTCCAAGCAGGACCCTGTGTATATCCCGCCCATCTCCTGCAACAAGGACCTG
TCCTTTGGCCACTCTAGGGCCAGCTCCAAGATCTGCAGTGAGGACATCGAGTGCAGTGGGCTGACCATCCCCAAGGC

TGTGCAGTACCTGAGCTCCCAGGATGAGAAGTACCAGGCCATTGGGGCCTATTACATCCAGCATACCTGCTTCCAGG
ATGAATCTGCCAAGCAACAGGTCTATCAGCTGGGAGGCATCTGCAAGCTGGTGGACCTCCTCCGCAGCCCCAACCAG
AACGTCCAGCAGGCCGCGGCAGGGGCCCTGCGCAACCTGGTGTTCAGGAGCACCACCAACAAGCTGGAGACCCGGAG
GCAGAATGGGATCCGCGAGGCAGTCAGCCTCCTGAGGAGAACCGGGAACGCCGAGATCCAGAAGCAGCTGACTGGGC
TGCTCTGGAACCTGTCTTCCACTGACGAGCTGAAGGAGGAACTCATTGCCGACGCCCTGCCTGTTCTGGCCGACCGC
GTCATCATTCCCTTCTCTGGCTGGTGCGATGGCAATAGCAACATGTCCCGGGAAGTGGTGGACCCTGAGGTCTTCTT
CAATGCCACAGGCTGCTTGAGGAACCTGAGCTCGGCCGATGCAGGCCGCCAGACCATGCGTAACTACTCAGGGCTCA
TTGATTCCCTCATGGCCTATGTCCAGAACTGTGTAGCGGCCAGCCGCTGTGACGACAAGTCTGTGGAAAACTGCATG
TGTGTTCTGCACAACCTCTCCTACCGCCTGGACGCCGAGGTGCCCACCCGCTACCGCCAGCTGGAGTATAACGCCCG
CAACGCCTACACCGAGAAGTCCTCCACTGGCTGCTTCAGCAACAAGAGCGACAAGATGATGAACAACAACTATGACT
GCCCCCTGCCTGAGGAAGAGACCAACCCCAAGGGCAGCGGCTGGTTGTACCATTCAGATGCCATCCGCACCTACCTG
AACCTCATGGGCAAGAGCAAGAAAGATGCTACCCTGGAGGCCTGTGCTGGTGCCCTGCAGAACCTGACAGCCAGCAA
GGGGCTGATGTCCAGTGGCATGAGCCAGTTGATTGGGCTGAAGGAAAAGGGCCTGCCACAAATTGCCCGCCTCCTGC
AATCTGGCAACTCTGATGTGGTGCGGTCCGGAGCCTCCCTCCTGAGCAACATGTCCCGCCACCCTCTGCTGCACAGA
GTGATGGGGAACCAGGTGTTCCCGGAGGTGACCAGGCTCCTCACCAGCCACACTGGCAATACCAGCAACTCCGAAGA
CATCTTGTCCTCGGCCTGCTACACTGTGAGGAACCTGATGGCCTCGCAGCCACAACTGGCCAAGCAGTACTTCTCCA
GCAGCATGCTCAACAACATCATCAACCTGTGCCGAAGCAGTGCCTCACCCAAGGCCGCAGAAGCTGCCCGGCTTCTC
CTGTCTGACATGTGGTCCAGCAAGGAACTGCAGGGTGTCCTCAGACAGCAAGGTTTCGATAGGAACATGCTGGGAAC
CTTAGCTGGGGCCAACAGCCTCAGGAACTTCACCTCCCGATTCTAAGAAGAGACTGTCCAAGCAAGTTAGGCTTGCA
GGAAGATATGACCCAGCTGAGAAGCCCTCAGGCCTCGCTGGATGGGGTTTTCTGTCCATCCTGTGCAGTATTTGGGA
AAGTTCACAAGAAACTGAGAAGAAACCTAAAAACTGTGGATAGTGGAAAGATTTTTAGATTTTTTTTTCCTTGGGG
AAACTGGCAGGCAATGGGGGTTAGGGAGGTTGGGGCGGGGGGGGCTTTCTTGAGTTAAAGGGGCTTATATGTGATGT
CAATATTTCTTCCTCTGAGAAATGGTATATATATGTGTCTAATGTAAGTGTGTGCATGCATGTGCGCGTGCATGTGT
GTGTGTGTGAGTGTCTTAAAGCATAACCACAAACTGCAAAAAGCTAGGTAAGCTATTTTGTTGCAGCTCATAAGGTG
GTGAAAAGGACTCTCCTGTGTTTCTTACTCATAGGCAAGGACAACATGTGCTTTTTGGTGAGCTGCTCATAATTCCT
GAAATGTGTGGTGCCAGGGCAAGGGGGCCATCACTGCAGTCAGGCCCTCAGAGGAGTCCTGCAGGCTTCCTACCAGT
GGTCTCCAAGGGTGCAGGAGTAACTGGGGCTGGGCCAGCCTCCCCCCTTACAAGGCTGCTTTCCAGGAAGGGAGGTC
TGGTGTATCTCATGGGAGAATCTGGGGTGTCTGTAGTGTCACCCCTCCAGCAGCGCCACAAGGACTGAGGTTGGGTA
GGTGTGAGGTTCCAGAGGACAGCAGGACACTCTCGCATACTTTGCCAAATGAGGCCTGCTCAGAGGAGTAGGAGCTG
AAAGATGGTGCCTTCCACCCTCTTGGGCTGTGTGCCCATCAGAGCAGGCTCAGCCTGCAAAGGCCCTGCATTCAGAG
GTCTTGTAATCTACTTGTTGCAGGAGAAAGAAGGTAAAAAATGATTTTTTTAAGAAAAGCTATTTTATTGCAGCTCT
TTCCCAAGAGCTGTTCTGGGAATGGCTGGTCTTCATATTCCCAGTGGAGAGGGGAACAAGTGGGGCTGGGCATATAC
CTATTCCGGCTTCTAGTGGGATGGAGTTGGGGTATAGAAATTAACCAGGAAGATGTTTCCACCAAGCCTGCTGTGAG
TCAATTGAGGGAGTGTTTGGGGTCCCAGGAGACTTGGACGGGGGGAGTTTGGGTAGACTAGGAAAGGAAAGTGCCAT
ATCAGGGTACCGGTACCGGCAAGCTCACATCTCAGCCAGGGGCCATGCCCCACTTCCCCTGACCCCAGCTGTCTTGT
CTCCACTCTGTGAAACCCACAGGGGATGTGATAAACAGGGCTATTAGGGGTATCAGCCACGTCGAGCCCCCAGACTC
TGTGCACTTCAGACCAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTATGAGAAAACCTGTGTGGACATCCCTTGGTGT
ACACTAAGACAGAGCAGAGCCCAGCGCTCCCAAGCCTTCCTCCTTCCAGCTTCTACCTCCATGCTAGCATTGCTGGT

GTTAGAGAGGAATTAACTTCCTGGTCTGTGCCCTTCTCTAGAAGAATATAAGATGCTCCTCCTCCTCACCCCTTCTC
AGCCTCCTCCCAAGTCTTCCTCTTCTGCACCACCCCCGAGTCCAAACCCACCTCTTGCCCCAGCATTCAGGCTGGAA
AACACTGATGTGGACTCAGTATGACAACTGAGATGGGGGAAGCCAGACATGTGAGGACGCTGTCCTCCGAGAGGTGT
CCCCGGCTGTTAGCCAGCTGTGCTGTGGTGCTGTGGGTCTGTCATACCCTCCCTTGCTTCTGTTCACACTGGGAGGC
CCACTCCTGGCTCACCTCTCCCTCTCAGGGACCCACGTGGGAGCCTGGATCCCTGGACTGTCCTGGGCATAGGTTTC
AGGGGCCTCCTTTGTTGTCATCAGAACCCAGAGGAATTCTTCTCCTAAAAAATACGTATGGCATACCAATCTGTGCG
GGGCAGTGTCCTAAGCACTTAGACTACATCAGGGAAGAACACAGACCACATCCCTGTCCTCATGCGGCTTATGTTTT
CTGGAGGAAAGTGGAGACACAAGTCCTTGGCTTTAGGGCTCCCCCGGCTGGGGGCTGTGCAGTCCGGTCAGGGCGGG
AGGGGAAATGCACCGCTGCATGTGAACCTTACCAGCCCAGGCGGATGCCCCTTCCCCTTAGCACTACCCTGGCCTCC
TGCATCCCCTCGCCTCATGTTCCTCCCACCTTCAAAGAATGAAGAGCCCCATGGGCCCAGCCCCTGCCCTGGGAACC
AGGCAGCCTTCCAGACCTCAGGGGCTGAGGCAGACTATTAGGGCAGGGCTGACTTTGGTGACACTGCCCATTCCCTC
TCAGGCCAGCTCAGGTCACCCGGGCCTCTGACCCAGGCCTGTCACTTTGAGAGGGGCAAAACTGAGAGGGGCTTTTC
CTAGAGAAAGAGAACAAGGAGCTTGCCAGGCTTCATGTAGCCGACACACGTCTCAGGATTTTAAGTCCACATTGGCC
TCACACTACCAGGGCCAATGCCCAAAATAAGGAGTTCCAATTTGGGGCCAAATGAGGAAGGACACAGACTCTGCCCT
GGGATCTCCTGTGCTAGCGGCCAATGACAAATCCAGTCATTGGCCACCAGCCACCTCTGCAGTGGGGACCACACTAG
CAGCCCTGACTCCACACTCCTCCTGGGGACCCAAGAGGCAGTGTTGCTGTCTGCGTGTCCACCTTGGAATCTGGCTG
AACTGGCTGGGAGGACCAAGACTGCGGCTGGGGTGGGCAGGGAAGGGAAGCCGGGGGCTGCTGTGAGGGATCTTGGA
GCTTCCCTGTAGCCCACCTTCCCCTTGCTTCATGTTTGTAGAGGAACCTTGTGCCGGCCAGGCCCAGTTTCCTTGTG
TGATACACTAATGTATTTGCTTTTTTTGGAAATAGAGAAAATCAATAAATTGCTAGTGTTTCTTTGAACTTTTTCGA

SEQ ID NO:1602

>HSB6PR # transcript_5 #len 5453 (Includes node 29 - TAA seg 34)

GGGGTGGTGCAGGGCAGGGGTGGTATATCCTGTCTGACGGAGGGCGGGCCTCGCCAGTGCCAGAGAGGGACGAACCA
GGGTGGAAGCGCCAGGAGCAGCTGCAGGGAGCCCTCACGCGGACCTCGCACTCTATGGCCGTAGGGAGCCGCTGAGA
GCGAGAAGAGCACGCTCCTGCCCGCCCGCTGCACCGCACCTCGCCTCGCCTCTCTGCTCTCCTAGGCCCCGGCCGCG
CGCCACCCGCCTCCCGCCACCATGAACCACTCGCCGCTCAAGACCGCCTTGGCGTACGAATGCTTCCAGGACCAGGA
CAACTCCACGTTGGCTTTGCCGTCGGACCAAAAGATGAAAACAGGCACGTCTGGCAGGCAGCGCGTGCAGGAGCAGG
TGATGATGACCGTCAAGCGGCAGAAGTCCAAGTCTTCCCAGTCGTCCACCCTGAGCCACTCCAATCGAGGTTCCATG
TATGATGGCTTGGCTGACAATTACAACTATGGGACCACCAGCAGGAGCAGCTACTACTCCAAGTTCCAGGCAGGGAA
TGGCTCATGGGGATATCCGATCTACAATGGAACCCTCAAGCGGGAGCCTGACAACAGGCGCTTCAGCTCCTACAGCC
AGATGGAGAACTGGAGCCGGCACTACCCCCGGGGCAGCTGTAACACCACCGGCGCAGGCAGCGACATCTGCTTCATG
CAGAAAATCAAGGCGAGCCGCAGTGAGCCCGACCTCTACTGTGACCCACGGGGCACCCTGCGCAAGGGCACGCTGGG
CAGCAAGGGCCAGAAGACCACCCAGAACCGCTACAGCTTTTACAGCACCTGCAGTGGTCAGAAGGCCATAAAGAAGT
GCCCTGTGCGCCCGCCCTCTTGTGCCTCCAAGCAGGACCCTGTGTATATCCCGCCCATCTCCTGCAACAAGGACCTG
TCCTTTGGCCACTCTAGGGCCAGCTCCAAGATCTGCAGTGAGGACATCGAGTGCAGTGGGCTGACCATCCCCAAGGC
TGTGCAGTACCTGAGCTCCCAGGATGAGAAGTACCAGGCCATTGGGGCCTATTACATCCAGCATACCTGCTTCCAGG
ATGAATCTGCCAAGCAACAGGTCTATCAGCTGGGAGGCATCTGCAAGCTGGTGGACCTCCTCCGCAGCCCCAACCAG
AACGTCCAGCAGGCCGCGGCAGGGGCCCTGCGCAACCTGGTGTTCAGGAGCACCACCAACAAGCTGGAGACCCGGAG

GCAGAATGGGATCCGCGAGGCAGTCAGCCTCCTGAGGAGAACCGGGAACGCCGAGATCCAGAAGCAGCTGACTGGGC
TGCTCTGGAACCTGTCTTCCACTGACGAGCTGAAGGAGGAACTCATTGCCGACGCCCTGCCTGTTCTGGCCGACCGC
GTCATCATTCCCTTCTCTGGCTGGTGCGATGGCAATAGCAACATGTCCCGGGAAGTGGTGGACCCTGAGGTCTTCTT
CAATGCCACAGGCTGCTTGAGAAAGAGACTGGGCATGCGGGAGCTTCTGGCTCTTGTTCCGCAAAGGGCCACTAGTA
GCAGGGTGAACCTGAGCTCGGCCGATGCAGGCCGCCAGACCATGCGTAACTACTCAGGGCTCATTGATTCCCTCATG
GCCTATGTCCAGAACTGTGTAGCGGCCAGCCGCTGTGACGACAAGTCTGTGGAAAACTGCATGTGTGTTCTGCACAA
CCTCTCCTACCGCCTGGACGCCGAGGTGCCCACCCGCTACCGCCAGCTGGAGTATAACGCCCGCAACGCCTACACCG
AGAAGTCCTCCACTGGCTGCTTCAGCAACAAGAGCGACAAGATGATGAACAACAACTATGACTGCCCCCTGCCTGAG
GAAGAGACCAACCCCAAGGGCAGCGGCTGGTTGTACCATTCAGATGCCATCCGCACCTACCTGAACCTCATGGGCAA
GAGCAAGAAAGATGCTACCCTGGAGGCCTGTGCTGGTGCCCTGCAGAACCTGACAGCCAGCAAGGGGCTGATGTCCA
GTGGCATGAGCCAGTTGATTGGGCTGAAGGAAAAGGGCCTGCCACAAATTGCCCGCCTCCTGCAATCTGGCAACTCT
GATGTGGTGCGGTCCGGAGCCTCCCTCCTGAGCAACATGTCCCGCCACCCTCTGCTGCACAGAGTGATGGGGAACCA
GGTGTTCCCGGAGGTGACCAGGCTCCTCACCAGCCACACTGGCAATACCAGCAACTCCGAAGACATCTTGTCCTCGG
CCTGCTACACTGTGAGGAACCTGATGGCCTCGCAGCCACAACTGGCCAAGCAGTACTTCTCCAGCAGCATGCTCAAC
AACATCATCAACCTGTGCCGAAGCAGTGCCTCACCCAAGGCCGCAGAAGCTGCCCGGCTTCTCCTGTCTGACATGTG
GTCCAGCAAGGAACTGCAGGGTGTCCTCAGACAGCAAGGTTTCGATAGGAACATGCTGGGAACCTTAGCTGGGGCCA
ACAGCCTCAGGAACTTCACCTCCCGATTCTAAGAAGAGACTGTCCAAGCAAGTTAGGCTTGCAGGAAGATATGACCC
AGCTGAGAAGCCCTCAGGCCTCGCTGGATGGGGTTTTCTGTCCATCCTGTGCAGTATTTGGGAAAGTTCACAAGAAA
CTGAGAAGAAACCTAAAAACTGTGGATAGTGGAAAGATTTTTAGATTTTTTTTTTCCTTGGGGAAACTGGCAGGCAA
TGGGGGTTAGGGAGGTTGGGGCGGGGGGGGCTTTCTTGAGTTAAAGGGGCTTATATGTGATGTCAATATTTCTTCCT
CTGAGAAATGGTATATATATGTGTCTAATGTAAGTGTGTGCATGCATGTGCGCGTGCATGTGTGTGTGTGTGAGTGT
CTTAAAGCATAACCACAAACTGCAAAAAGCTAGGTAAGCTATTTTGTTGCAGCTCATAAGGTGGTGAAAAGGACTCT
CCTGTGTTTCTTACTCATAGGCAAGGACAACATGTGCTTTTTGGTGAGCTGCTCATAATTCCTGAAATGTGTGGTGC
CAGGGCAAGGGGGCCATCACTGCAGTCAGGCCCTCAGAGGAGTCCTGCAGGCTTCCTACCAGTGGTCTCCAAGGGTG
CAGGAGTAACTGGGGCTGGGCCAGCCTCCCCCCTTACAAGGCTGCTTTCCAGGAAGGGAGGTCTGGTGTATCTCATG
GGAGAATCTGGGGTGTCTGTAGTGTCACCCCTCCAGCAGCGCCACAAGGACTGAGGTTGGGTAGGTGTGAGGTTCCA
GAGGACAGCAGGACACTCTCGCATACTTTGCCAAATGAGGCCTGCTCAGAGGAGTAGGAGCTGAAAGATGGTGCCTT
CCACCCTCTTGGGCTGTGTGCCCATCAGAGCAGGCTCAGCCTGCAAAGGCCCTGCATTCAGAGGTCTTGTAATCTAC
TTGTTGCAGGAGAAAGAAGGTAAAAAATGATTTTTTTAAGAAAAGCTATTTTATTGCAGCTCTTTCCCAAGAGCTGT
TCTGGGAATGGCTGGTCTTCATATTCCCAGTGGAGAGGGGAACAAGTGGGGCTGGGCATATACCTATTCCGGCTTCT
AGTGGGATGGAGTTGGGGTATAGAAATTAACCAGGAAGATGTTTCCACCAAGCCTGCTGTGAGTCAATTGAGGGAGT
GTTTGGGGTCCCAGGAGACTTGGACGGGGGGAGTTTGGGTAGACTAGGAAAGGAAAGTGCCATATCAGGGTACCGGT
ACCGGCAAGCTCACATCTCAGCCAGGGGCCATGCCCCACTTCCCCTGACCCCAGCTGTCTTGTCTCCACTCTGTGAA
ACCCACAGGGGATGTGATAAACAGGGCTATTAGGGGTATCAGCCACGTCGAGCCCCCAGACTCTGTGCACTTCAGAC
CAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTATGAGAAAACCTGTGTGGACATCCCTTGGTGTACACTAAGACAGAG
CAGAGCCCAGCGCTCCCAAGCCTTCCTCCTTCCAGCTTCTACCTCCATGCTAGCATTGCTGGTGTTAGAGAGGAATT
AACTTCCTGGTCTGTGCCCTTCTCTAGAAGAATATAAGATGCTCCTCCTCCTCACCCCTTCTCAGCCTCCTCCCAAG
TCTTCCTCTTCTGCACCACCCCCGAGTCCAAACCCACCTCTTGCCCCAGCATTCAGGCTGGAAAACACTGATGTGGA

CTCAGTATGACAACTGAGATGGGGGAAGCCAGACATGTGAGGACGCTGTCCTCCGAGAGGTGTCCCCGGCTGTTAGC
CAGCTGTGCTGTGGTGCTGTGGGTCTGTCATACCCTCCCTTGCTTCTGTTCACACTGGGAGGCCCACTCCTGGCTCA
CCTCTCCCTCTCAGGGACCCACGTGGGAGCCTGGATCCCTGGACTGTCCTGGGCATAGGTTTCAGGGGCCTCCTTTG
TTGTCATCAGAACCCAGAGGAATTCTTCTCCTAAAAAATACGTATGGCATACCAATCTGTGCGGGGCAGTGTCCTAA
GCACTTAGACTACATCAGGGAAGAACACAGACCACATCCCTGTCCTCATGCGGCTTATGTTTTCTGGAGGAAAGTGG
AGACACAAGTCCTTGGCTTTAGGGCTCCCCCGGCTGGGGGCTGTGCAGTCCGGTCAGGGCGGGAGGGGAAATGCACC
GCTGCATGTGAACCTTACCAGCCCAGGCGGATGCCCCTTCCCCTTAGCACTACCCTGGCCTCCTGCATCCCCTCGCC
TCATGTTCCTCCCACCTTCAAAGAATGAAGAGCCCCATGGGCCCAGCCCCTGCCCTGGGAACCAGGCAGCCTTCCAG
ACCTCAGGGGCTGAGGCAGACTATTAGGGCAGGGCTGACTTTGGTGACACTGCCCATTCCCTCTCAGGCCAGCTCAG
GTCACCCGGGCCTCTGACCCAGGCCTGTCACTTTGAGAGGGGCAAAACTGAGAGGGGCTTTTCCTAGAGAAAGAGAA
CAAGGAGCTTGCCAGGCTTCATGTAGCCGACACACGTCTCAGGATTTTAAGTCCACATTGGCCTCACACTACCAGGG
CCAATGCCCAAAATAAGGAGTTCCAATTTGGGGCCAAATGAGGAAGGACACAGACTCTGCCCTGGGATCTCCTGTGC
TAGCGGCCAATGACAAATCCAGTCATTGGCCACCAGCCACCTCTGCAGTGGGGACCACACTAGCAGCCCTGACTCCA
CACTCCTCCTGGGGACCCAAGAGGCAGTGTTGCTGTCTGCGTGTCCACCTTGGAATCTGGCTGAACTGGCTGGGAGG
ACCAAGACTGCGGCTGGGGTGGGCAGGGAAGGGAAGCCGGGGGCTGCTGTGAGGGATCTTGGAGCTTCCCTGTAGCC
CACCTTCCCCTTGCTTCATGTTTGTAGAGGAACCTTGTGCCGGCCAGGCCCAGTTTCCTTGTGTGATACACTAATGT
ATTTGCTTTTTTTGGAAATAGAGAAAATCAATAAATTGCTAGTGTTTCTTTGAACTTTTTCGA

SEQ ID NO:1603
>HSB6PR # transcript_6 #len 4543 (Includes node 29 - TAA seg 34 ; node 33 - TAA seg 8)
TTCCATTTTCCTAATACCAGACACAAACAAGAGAGATGGCAAGGATCTGCAGTGAGGACATCGAGTGCAGTGGGCTG
ACCATCCCCAAGGCTGTGCAGTACCTGAGCTCCCAGGATGAGAAGTACCAGGCCATTGGGGCCTATTACATCCAGCA
TACCTGCTTCCAGGATGAATCTGCCAAGCAACAGGTCTATCAGCTGGGAGGCATCTGCAAGCTGGTGGACCTCCTCC
GCAGCCCCAACCAGAACGTCCAGCAGGCCGCGGCAGGGGCCCTGCGCAACCTGGTGTTCAGGAGCACCACCAACAAG
CTGGAGACCCGGAGGCAGAATGGGATCCGCGAGGCAGTCAGCCTCCTGAGGAGAACCGGGAACGCCGAGATCCAGAA
GCAGCTGACTGGGCTGCTCTGGAACCTGTCTTCCACTGACGAGCTGAAGGAGGAACTCATTGCCGACGCCCTGCCTG
TTCTGGCCGACCGCGTCATCATTCCCTTCTCTGGCTGGTGCGATGGCAATAGCAACATGTCCCGGGAAGTGGTGGAC
CCTGAGGTCTTCTTCAATGCCACAGGCTGCTTGAGAAAGAGACTGGGCATGCGGGAGCTTCTGGCTCTTGTTCCGCA
AAGGGCCACTAGTAGCAGGGTGAACCTGAGCTCGGCCGATGCAGGCCGCCAGACCATGCGTAACTACTCAGGGCTCA
TTGATTCCCTCATGGCCTATGTCCAGAACTGTGTAGCGGCCAGCCGCTGTGACGACAAGTCTGTGGAAAACTGCATG
TGTGTTCTGCACAACCTCTCCTACCGCCTGGACGCCGAGGTGCCCACCCGCTACCGCCAGCTGGAGTATAACGCCCG
CAACGCCTACACCGAGAAGTCCTCCACTGGCTGCTTCAGCAACAAGAGCGACAAGATGATGAACAACAACTATGACT
GCCCCCTGCCTGAGGAAGAGACCAACCCCAAGGGCAGCGGCTGGTTGTACCATTCAGATGCCATCCGCACCTACCTG
AACCTCATGGGCAAGAGCAAGAAAGATGCTACCCTGGAGGCCTGTGCTGGTGCCCTGCAGAACCTGACAGCCAGCAA
GGGGCTGATGTCCAGTGGCATGAGCCAGTTGATTGGGCTGAAGGAAAAGGGCCTGCCACAAATTGCCCGCCTCCTGC
AATCTGGCAACTCTGATGTGGTGCGGTCCGGAGCCTCCCTCCTGAGCAACATGTCCCGCCACCCTCTGCTGCACAGA
GTGATGGGGAACCAGGTGTTCCCGGAGGTGACCAGGCTCCTCACCAGCCACACTGGCAATACCAGCAACTCCGAAGA

CATCTTGTCCTCGGCCTGCTACACTGTGAGGAACCTGATGGCCTCGCAGCCACAACTGGCCAAGCAGTACTTCTCCA
GCAGCATGCTCAACAACATCATCAACCTGTGCCGAAGCAGTGCCTCACCCAAGGCCGCAGAAGCTGCCCGGCTTCTC
CTGTCTGACATGTGGTCCAGCAAGGAACTGCAGGGTGTCCTCAGACAGCAAGGTTTCGATAGGAACATGCTGGGAAC
CTTAGCTGGGGCCAACAGCCTCAGGAACTTCACCTCCCGATTCTAAGAAGAGACTGTCCAAGCAAGTTAGGCTTGCA
GGAAGATATGACCCAGCTGAGAAGCCCTCAGGCCTCGCTGGATGGGGTTTTCTGTCCATCCTGTGCAGTATTTGGGA
AAGTTCACAAGAAACTGAGAAGAAACCTAAAAACTGTGGATAGTGGAAAGATTTTTAGATTTTTTTTTTCCTTGGGG
AAACTGGCAGGCAATGGGGGTTAGGGAGGTTGGGGCGGGGGGGGCTTTCTTGAGTTAAAGGGGCTTATATGTGATGT
CAATATTTCTTCCTCTGAGAAATGGTATATATATGTGTCTAATGTAAGTGTGTGCATGCATGTGCGCGTGCATGTGT
GTGTGTGTGAGTGTCTTAAAGCATAACCACAAACTGCAAAAAGCTAGGTAAGCTATTTTGTTGCAGCTCATAAGGTG
GTGAAAAGGACTCTCCTGTGTTTCTTACTCATAGGCAAGGACAACATGTGCTTTTTGGTGAGCTGCTCATAATTCCT
GAAATGTGTGGTGCCAGGGCAAGGGGGCCATCACTGCAGTCAGGCCCTCAGAGGAGTCCTGCAGGCTTCCTACCAGT
GGTCTCCAAGGGTGCAGGAGTAACTGGGGCTGGGCCAGCCTCCCCCCTTACAAGGCTGCTTTCCAGGAAGGGAGGTC
TGGTGTATCTCATGGGAGAATCTGGGGTGTCTGTAGTGTCACCCCTCCAGCAGCGCCACAAGGACTGAGGTTGGGTA
GGTGTGAGGTTCCAGAGGACAGCAGGACACTCTCGCATACTTTGCCAAATGAGGCCTGCTCAGAGGAGTAGGAGCTG
AAAGATGGTGCCTTCCACCCTCTTGGGCTGTGTGCCCATCAGAGCAGGCTCAGCCTGCAAAGGCCCTGCATTCAGAG
GTCTTGTAATCTACTTGTTGCAGGAGAAAGAAGGTAAAAAATGATTTTTTTAAGAAAAGCTATTTTATTGCAGCTCT
TTCCCAAGAGCTGTTCTGGGAATGGCTGGTCTTCATATTCCCAGTGGAGAGGGGAACAAGTGGGGCTGGGCATATAC
CTATTCCGGCTTCTAGTGGGATGGAGTTGGGGTATAGAAATTAACCAGGAAGATGTTTCCACCAAGCCTGCTGTGAG
TCAATTGAGGGAGTGTTTGGGGTCCCAGGAGACTTGGACGGGGGGAGTTTGGGTAGACTAGGAAAGGAAAGTGCCAT
ATCAGGGTACCGGTACCGGCAAGCTCACATCTCAGCCAGGGGCCATGCCCCACTTCCCCTGACCCCAGCTGTCTTGT
CTCCACTCTGTGAAACCCACAGGGGATGTGATAAACAGGGCTATTAGGGGTATCAGCCACGTCGAGCCCCCAGACTC
TGTGCACTTCAGACCAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTATGAGAAAACCTGTGTGGACATCCCTTGGTGT
ACACTAAGACAGAGCAGAGCCCAGCGCTCCCAAGCCTTCCTCCTTCCAGCTTCTACCTCCATGCTAGCATTGCTGGT
GTTAGAGAGGAATTAACTTCCTGGTCTGTGCCCTTCTCTAGAAGAATATAAGATGCTCCTCCTCCTCACCCCTTCTC
AGCCTCCTCCCAAGTCTTCCTCTTCTGCACCACCCCCGAGTCCAAACCCACCTCTTGCCCCAGCATTCAGGCTGGAA
AACACTGATGTGGACTCAGTATGACAACTGAGATGGGGGAAGCCAGACATGTGAGGACGCTGTCCTCCGAGAGGTGT
CCCCGGCTGTTAGCCAGCTGTGCTGTGGTGCTGTGGGTCTGTCATACCCTCCCTTGCTTCTGTTCACACTGGGAGGC
CCACTCCTGGCTCACCTCTCCCTCTCAGGGACCCACGTGGGAGCCTGGATCCCTGGACTGTCCTGGGCATAGGTTTC
AGGGGCCTCCTTTGTTGTCATCAGAACCCAGAGGAATTCTTCTCCTAAAAAATACGTATGGCATACCAATCTGTGCG
GGGCAGTGTCCTAAGCACTTAGACTACATCAGGGAAGAACACAGACCACATCCCTGTCCTCATGCGGCTTATGTTTT
CTGGAGGAAAGTGGAGACACAAGTCCTTGGCTTTAGGGCTCCCCCGGCTGGGGGCTGTGCAGTCCGGTCAGGGCGGG
AGGGGAAATGCACCGCTGCATGTGAACCTTACCAGCCCAGGCGGATGCCCCTTCCCCTTAGCACTACCCTGGCCTCC
TGCATCCCCTCGCCTCATGTTCCTCCCACCTTCAAAGAATGAAGAGCCCCATGGGCCCAGCCCCTGCCCTGGGAACC
AGGCAGCCTTCCAGACCTCAGGGGCTGAGGCAGACTATTAGGGCAGGGCTGACTTTGGTGACACTGCCCATTCCCTC
TCAGGCCAGCTCAGGTCACCCGGGCCTCTGACCCAGGCCTGTCACTTTGAGAGGGGCAAAACTGAGAGGGGCTTTTC
CTAGAGAAAGAGAACAAGGAGCTTGCCAGGCTTCATGTAGCCGACACACGTCTCAGGATTTTAAGTCCACATTGGCC
TCACACTACCAGGGCCAATGCCCAAAATAAGGAGTTCCAATTTGGGGCCAAATGAGGAAGGACACAGACTCTGCCCT
GGGATCTCCTGTGCTAGCGGCCAATGACAAATCCAGTCATTGGCCACCAGCCACCTCTGCAGTGGGGACCACACTAG

CAGCCCTGACTCCACACTCCTCCTGGGGACCCAAGAGGCAGTGTTGCTGTCTGCGTGTCCACCTTGGAATCTGGCTG
AACTGGCTGGGAGGACCAAGACTGCGGCTGGGGTGGGCAGGGAAGGGAAGCCGGGGGCTGCTGTGAGGGATCTTGGA
GCTTCCCTGTAGCCCACCTTCCCCTTGCTTCATGTTTGTAGAGGAACCTTGTGCCGGCCAGGCCCAGTTTCCTTGTG
TGATACACTAATGTATTTGCTTTTTTTGGAAATAGAGAAAATCAATAAATTGCTAGTGTTTCTTTGAACTTTTTCGA

SEQ ID NO:1604
>HSB6PR_PEA_2_P1 # trn_0 #len 726
MNHSPLKTALAYECFQDQDNSTLALPSDQKMKTGTSGRQRVQEQVMMTVKRQKSKSSQSSTLSHSNRGSMYDGLADN
YNYGTTSRSSYYSKFQAGNGSWGYPIYNGTLKREPDNRRFSSYSQMENWSRHYPRGSCNTTGAGSDICFMQKIKASR
SEPDLYCDPRGTLRKGTLGSKGQKTTQNRYSFYSTCSGQKAIKKCPVRPPSCASKQDPVYIPPISCNKDLSFGHSRA
SSKICSEDIECSGLTIPKAVQYLSSQDEKYQAIGAYYIQHTCFQDESAKQQVYQLGGICKLVDLLRSPNQNVQQAAA
GALRNLVFRSTTNKLETRRQNGIREAVSLLRRTGNAEIQKQLTGLLWNLSSTDELKEELIADALPVLADRVIIPFSG
WCDGNSNMSREVVDPEVFFNATGCLRNLSSADAGRQTMRNYSGLIDSLMAYVQNCVAASRCDDKSVENCMCVLHNLS
YRLDAEVPTRYRQLEYNARNAYTEKSSTGCFSNKSDKMMNNNYDCPLPEEETNPKGSGWLYHSDAIRTYLNLMGKSK
KDATLEACAGALQNLTASKGLMSSGMSQLIGLKEKGLPQIARLLQSGNSDVVRSGASLLSNMSRHPLLHRVMGNQVF
PEVTRLLTSHTGNTSNSEDILSSACYTVRNLMASQPQLAKQYFSSSMLNNIINLCRSSASPKAAEAARLLLSDMWSS
KELQGVLRQQGFDRNMLGTLAGANSLRNFTSRF

SEQ ID NO:1605
>HSB6PR_PEA_2_P6 # trn_5 #len 747
MNHSPLKTALAYECFQDQDNSTLALPSDQKMKTGTSGRQRVQEQVMMTVKRQKSKSSQSSTLSHSNRGSMYDGLADN
YNYGTTSRSSYYSKFQAGNGSWGYPIYNGTLKREPDNRRFSSYSQMENWSRHYPRGSCNTTGAGSDICFMQKIKASR
SEPDLYCDPRGTLRKGTLGSKGQKTTQNRYSFYSTCSGQKAIKKCPVRPPSCASKQDPVYIPPISCNKDLSFGHSRA
SSKICSEDIECSGLTIPKAVQYLSSQDEKYQAIGAYYIQHTCFQDESAKQQVYQLGGICKLVDLLRSPNQNVQQAAA
GALRNLVFRSTTNKLETRRQNGIREAVSLLRRTGNAEIQKQLTGLLWNLSSTDELKEELIADALPVLADRVIIPFSG
WCDGNSNMSREVVDPEVFFNATGCLRKRLGMRELLALVPQRATSSRVNLSSADAGRQTMRNYSGLIDSLMAYVQNCV
AASRCDDKSVENCMCVLHNLSYRLDAEVPTRYRQLEYNARNAYTEKSSTGCFSNKSDKMMNNNYDCPLPEEETNPKG
SGWLYHSDAIRTYLNLMGKSKKDATLEACAGALQNLTASKGLMSSGMSQLIGLKEKGLPQIARLLQSGNSDVVRSGA
SLLSNMSRHPLLHRVMGNQVFPEVTRLLTSHTGNTSNSEDILSSACYTVRNLMASQPQLAKQYFSSSMLNNIINLCR
SSASPKAAEAARLLLSDMWSSKELQGVLRQQGFDRNMLGTLAGANSLRNFTSRF

SEQ ID NO:1606
>HSB6PR_PEA_2_P7 # trn_6 #len 516
MARICSEDIECSGLTIPKAVQYLSSQDEKYQAIGAYYIQHTCFQDESAKQQVYQLGGICKLVDLLRSPNQNVQQAAA
GALRNLVFRSTTNKLETRRQNGIREAVSLLRRTGNAEIQKQLTGLLWNLSSTDELKEELIADALPVLADRVIIPFSG
WCDGNSNMSREVVDPEVFFNATGCLRKRLGMRELLALVPQRATSSRVNLSSADAGRQTMRNYSGLIDSLMAYVQNCV
AASRCDDKSVENCMCVLHNLSYRLDAEVPTRYRQLEYNARNAYTEKSSTGCFSNKSDKMMNNNYDCPLPEEETNPKG
SGWLYHSDAIRTYLNLMGKSKKDATLEACAGALQNLTASKGLMSSGMSQLIGLKEKGLPQIARLLQSGNSDVVRSGA

SLLSNMSRHPLLHRVMGNQVFPEVTRLLTSHTGNTSNSEDILSSACYTVRNLMASQPQLAKQYFSSSMLNNIINLCR
SSASPKAAEAARLLLSDMWSSKELQGVLRQQGFDRNMLGTLAGANSLRNFTSRF

SEQ ID NO:1607
>HSB6PR # node_33 (TAA seg 8) #len 43
TTCCATTTTCCTAATACCAGACACAAACAAGAGAGATGGCAAG

SEQ ID NO:1608
>HSB6PR # node_29 (TAA seg 34) #len 1199
CTCTGAGAAATGGTATATATATGTGTCTAATGTAAGTGTGTGCATGCATGTGCGCGTGCATGTGTGTGTGTGTGAGT
GTCTTAAAGCATAACCACAAACTGCAAAAAGCTAGGTAAGCTATTTTGTTGCAGCTCATAAGGTGGTGAAAAGGACT
CTCCTGTGTTTCTTACTCATAGGCAAGGACAACATGTGCTTTTTGGTGAGCTGCTCATAATTCCTGAAATGTGTGGT
GCCAGGGCAAGGGGGCCATCACTGCAGTCAGGCCCTCAGAGGAGTCCTGCAGGCTTCCTACCAGTGGTCTCCAAGGG
TGCAGGAGTAACTGGGGCTGGGCCAGCCTCCCCCCTTACAAGGCTGCTTTCCAGGAAGGGAGGTCTGGTGTATCTCA
TGGGAGAATCTGGGGTGTCTGTAGTGTCACCCCTCCAGCAGCGCCACAAGGACTGAGGTTGGGTAGGTGTGAGGTTC
CAGAGGACAGCAGGACACTCTCGCATACTTTGCCAAATGAGGCCTGCTCAGAGGAGTAGGAGCTGAAAGATGGTGCC
TTCCACCCTCTTGGGCTGTGTGCCCATCAGAGCAGGCTCAGCCTGCAAAGGCCCTGCATTCAGAGGTCTTGTAATCT
ACTTGTTGCAGGAGAAAGAAGGTAAAAAATGATTTTTTTAAGAAAAGCTATTTTATTGCAGCTCTTTCCCAAGAGCT
GTTCTGGGAATGGCTGGTCTTCATATTCCCAGTGGAGAGGGGAACAAGTGGGGCTGGGCATATACCTATTCCGGCTT
CTAGTGGGATGGAGTTGGGGTATAGAAATTAACCAGGAAGATGTTTCCACCAAGCCTGCTGTGAGTCAATTGAGGGA
GTGTTTGGGGTCCCAGGAGACTTGGACGGGGGGAGTTTGGGTAGACTAGGAAAGGAAAGTGCCATATCAGGGTACCG
GTACCGGCAAGCTCACATCTCAGCCAGGGGCCATGCCCCACTTCCCCTGACCCCAGCTGTCTTGTCTCCACTCTGTG
AAACCCACAGGGGATGTGATAAACAGGGCTATTAGGGGTATCAGCCACGTCGAGCCCCCAGACTCTGTGCACTTCAG
ACCAGCAGCAGCAGGAGGGCTCCCGAGGGCCTTATGAGAAAACCTGTGTGGACATCCCTTGGTGTACACTAAGACAG
AGCAGAGCCCAGCGCTCCCAAGCCTTCCTCCTTCCAGCTTCTAC

SEQ ID NO:1609
>T86235 # transcript_31 #len 2871 (Includes node 39 - TAA seg 44; node 37 - TAA seg 42)
CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCACCATCCTGTCAGGTG
AGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAGGAAGTCAGGGAGGC
ACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGC
CAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTC
TGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGC

AGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAA
AGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAA
GAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAG
CCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAAC
TCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCA
GAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAG
GTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCC
TTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGT
TGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAA
CTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGG
ACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAG
AGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTA
CCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTC
TCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTG
AGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCC
CTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTG
CACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGA
TCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAG
GCAGGTAAGGAGTTGGCTGGGAAGGAGTGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGC
CCCTACCCCAGGCAATCTCATGCTTCCACACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCAGGAAGAGGGGAGGG
GCTGCACTTCCAGCCCTGTGCTCCTAATTGGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACAGTACATGGTGGAAG
TATAGGACCCCAGACCTCCCTCTAAATTTTCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCT
GAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTT
ACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGAT
GCCCTGGTGAGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGCCTGATGGGAGGTGGGGAACAGGGACAGG
GGGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTACAGTGCAAGGTTGGCTGAGCTGTGACAAG
GTCTTCTCTGTCTCCAGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCC
TGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGT
CCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1610
>T86235 # transcript_32 #len 2514 (Includes node 39 - TAA seg 44)
CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCACCATCCTGTCAGGTG
AGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAGGAAGTCAGGGAGGC

ACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGC
CAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTC
TGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGC
AGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAA
AGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAA
GAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAG
CCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAAC
TCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCA
GAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAG
GTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCC
TTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGT
TGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAA
CTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGG
ACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAG
AGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTA
CCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTC
TCGCCAGGAACAGCTTGAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCCCTTCAGCCCA
GCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTGCACCCTGGAA
CATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGATCTTCTCTTC
CCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAGGCAGGCCTCA
GCAATCTGGCCCCTCGAACCCTAGCCCTGAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAG
GCTCGTCTGGACGATGAGTGTGCCTTTTACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCC
TGTGGCTACATTACTCGAATGGCAGGATGCCCTGGTGAGACTCCAACCCACAGCCCAGCTGTGGCTGCACAGTGAGC
CTGATGGGAGGTGGGGAACAGGGACAGGGGGCCACCTGGGCTTCTTCACAGAGAGGTCAGCAGGAAGGCTTGGCTAC
AGTGCAAGGTTGGCTGAGCTGTGACAAGGTCTTCTCTGTCTCCAGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCA
GGGCTCTCCATGATGAGACAACCACTCCTGCCCTGCCGTACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCC
CATGGGACTGGGAGCCGCCCACTTTTGTCCTCAATAAAGTTTCTAAAGTA

SEQ ID NO:1611

>T86235 # transcript_33 #len 2706 (Includes node 37 - TAA seg 42)

CTCCAGCAGCACCCGAGAGGGTCAGGAGAAAAGCGGAGGAAGCTGGGTAGGCCCTGAGGGGCCTCGGTAAGCCATCA
TGACCACCCGGCAAGCCACGAAGGATCCCCTCCTCCGGGGGTGTATCTCCTACCCCTAGCAAGATTCCGGTACGCTCT
CAGAAACGCACGCCTTTCCCCACTGTTACATCGTGCGCCGTGGACCAGGAGAACCAAGATCCAAGGAGATGGGTGCA
GAAACCACCGCTCAATATTCAACGCCCCCTCGTTGATTCAGCAGGCCCCAGGCCGAAAGCCAGGCACCAGGCAGAGA
CATCACAAAGATTGAGGCTCCAGGGACCATAGAGTTTGTGGCTGACCCTGCAGCCCTGGCCACCATCCTGTCAGGTG
AGGGTGTGAAGAGCTGTCACCTGGGGCGCCAGCCTAGTCTGGCTAAAAGAGTACTGGTTCGAGGAAGTCAGGGAGGC
ACCACCCAGAGGGTCCAGGGTGTTCGGGCCTCTGCATATTTGGCCCCCAGAACCCCCACCCACCGACTGGACCCTGC
CAGGGCTTCCTGCTTCTCTAGGCTGGAGGGACCAGGACCTCGAGGCCGGACATTGTGCCCCCAGAGGCTACAGGCTC

TGATTTCACCTTCAGGACCTTCCTTTCACCCTTCCACTCGCCCCAGTTTCCAGGAGCTAAGAAGGGAGACAGCTGGC
AGCAGCCGGACTTCAGTGAGCCAGGCCTCAGGATTGCTCCTGGAGACCCCAGTCCAGCCTGCTTTCTCTCTTCCTAA
AGGAGAACGCGAGGTTGTCACTCACTCAGATGAAGGAGGTGTGGCCTCTCTTGGTCTGGCCCAGCGAGTACCATTAA
GAGAAAACCGAGAAATGTCACATACCAGGGACAGCCATGACTCCCACCTGATGCCCTCCCCTGCCCCTGTGGCCCAG
CCCTTGCCTGGCCATGTGGTGCCATGTCCATCACCCTTTGGACGGGCTCAGCGTGTACCCTCCCCAGGCCCTCCAAC
TCTGACCTCATATTCAGTGTTGCGGCGTCTCACCGTTCAACCTAAAACCCGGTTCACACCCATGCCATCAACCCCCA
GAGTTCAGCAGGCCCAGTGGCTGCGTGGTGTCTCCCCTCAGTCCTGCTCTGAAGATCCTGCCCTGCCCTGGGAGCAG
GTTGCCGTCCGGTTGTTTGACCAGGAGAGTTGTATAAGGTCACTGGAGGGTTCTGGGAAACCACCGGTGGCCACTCC
TTCTGGACCCCACTCTAACAGAACCCCCAGCCTCCAGGAGGTGAAGATTCAACGCATCGGTATCCTGCAACAGCTGT
TGAGACAGGAAGTAGAGGGGCTGGTAGGGGGCCAGTGTGTCCCTCTTAATGGAGGCTCTTCTCTGGATATGGTTGAA
CTTCAGCCCCTGCTGACTGAGATTTCTAGAACTCTGAATGCCACAGAGCATAACTCTGGGACTTCCCACCTTCCTGG
ACTGTTAAAACACTCAGGGCTGCCAAAGCCCTGTCTTCCAGAGGAGTGCGGGGAACCACAGCCCTGCCCTCCGGCAG
AGCCTGGGCCCCCAGAGGCCTTCTGTAGGAGTGAGCCTGAGATACCAGAGCCCTCCCTCCAGGAACAGCTTGAAGTA
CCAGAGCCCTACCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTGCTGTAGGAGTGAGCCTGAGATACCGGAGTCCTC
TCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCAGGCCCCTAGAGTCCTACTGTAGGATTG
AGCCTGAGATACCGGAGTCCTCTCGCCAGGAACAGCTTGAGGTACCTGAGCCCTGCCCTCCAGCAGAACCCGGGCCC
CTTCAGCCCAGCACCCAGGGGCAGTCTGGACCCCCAGGGCCCTGCCCTAGGGTAGAGCTGGGGGCATCAGAGCCCTG
CACCCTGGAACATAGAAGTCTAGAGTCCAGTCTACCACCCTGCTGCAGTCAGTGGGCTCCAGCAACCACCAGCCTGA
TCTTCTCTTCCCAACACCCGCTTTGTGCCAGCCCCCCTATCTGCTCACTCCAGTCTTTGAGACCCCCAGCAGGCCAG
GCAGGTAAGGAGTTGGCTGGGAAGGAGTGTGAACACAAGAGGTCCTCACCTCACTGTGAGCTGCACACCTGCCCTGC
CCCTACCCCAGGCAATCTCATGCTTCCACACCTTCCACCCTGGCCCAGCCTGGCTCTCCCTCAGGAAGAGGGGAGGG
GCTGCACTTCCAGCCCTGTGCTCCTAATTGGCTTGGCCGTTGGTGGGGGAGGAGGAGAGGACAGTACATGGTGGAAG
TATAGGACCCCAGACCTCCCTCTAAATTTTCCATGCCCCTCAGGCCTCAGCAATCTGGCCCCTCGAACCCTAGCCCT
GAGGGAGCGCCTCAAATCGTGTTTAACCGCCATCCACTGCTTCCACGAGGCTCGTCTGGACGATGAGTGTGCCTTTT
ACACCAGCCGAGCCCCTCCCTCAGGCCCCACCCGGGTCTGCACCAACCCTGTGGCTACATTACTCGAATGGCAGGAT
GCCCTGTGTTTCATTCCAGTTGGTTCTGCTGCCCCCCAGGGCTCTCCATGATGAGACAACCACTCCTGCCCTGCCGT
ACTTCTTCCTTTTAGCCCTTATTTATTGTCGGTCTGCCCATGGGACTGGGAGCCGCCCACTTTTGTCCTCAATAAAG
TTTCTAAAGTA

SEQ ID NO:1612
>T86235_PEA_13_P25 # trn_31, 32, 33 #len 87
MTTRQATKDPLLRGVSPTPSKIPVRSQKRTPFPTVTSCAVDQENQDPRRWVQKPPLNIQRPLVDSAGPRPKARHQAE
TSQRLRLQGP

SEQ ID NO:1613
>Unique aa coded by T31,32,33 [found in T86235_PEA_13_P25]
RLQGP

SEQ ID NO:1614
>Forward primer -
TTTTCAAATGGGTAGGGACCATC

SEQ ID NO:1615
>Reverse primer -
TGAGTTTCTCTGGGACCCGGA

SEQ ID NO:1616
>Amplicon seq:
TTTTCAAATGGGTAGGGACCATCCATGGAGCAGCTGGAACAGCAGTGGGGAGCATCAGAACCAGCTCAACAGTTTGT
CTACTGTCCGGTCCCAGAGAAACTCA

SEQ ID NO:1617
>Forward primer -
TGTTTCTCCAAATGCCAGAACC

SEQ ID NO:1618
>Reverse primer -
GGCTGGTGACCTGCTTTGA

SEQ ID NO:1619
>Amplicon seq:
TGTTTCTCCAAATGCCAGAACCCAACATTGATAGTCCCTTGAACACACATGCTGCCGAGCTCTGGAAAAACCCCACA
GCTTTTAAGAAGTACCTGCAAGAAACCTACTCAAAGCAGGTCACCAGCC

SEQ ID NO:1620
>Forward primer -
GAGGCGAGGAGTGTGGCAC

SEQ ID NO:1621
>Reverse primer -
GCTGCGATGGGCACGTT

SEQ ID NO:1622
>Amplicon:
GAGGCGAGGAGTGTGGCACTTTGGCGGGGAAGGGGCGGCTCAGCCCTCGGGCCCTCGCCCGCCCTCTCCGGGTCTGG
AGCGTCTCCTCGCGCCATCCCTGCACCGCCAGGGGGAACGTGCCCATCGCAGC

SEQ ID NO:1623
>Forward primer -
CCTAGCAGCTCGTCTCCAGC

SEQ ID NO:1624
>Reverse primer -
GGTGTCCTCGCCGAACCT

SEQ ID NO:1625
>Amplicon
CCTAGCAGCTCGTCTCCAGCTAAAAATCCTCCAGAGCCACCGGCGAACACTGCTAGGCACAGCCTGGTGCCAAGCTA
CGAAGCGCCCGCCGCCGCCGTGCGCAGGTTCGGCGAGGACACC

SEQ ID No.: 1626
>T23580_T10
TTTTTGTTTGTGTACTTAATAAAGGGTAAATATGTCATGTTTGTTTGGAACAGTCATGGTTAATATCTATGTTGTCC
CAGTATATCTATTAATAGAACTCTCTTTCACTCTCAACAGCGTCCTAGTCCGGATGACAAATTATATGGTTATCTCT
CAGTAAAGGGTCTTTTTTTAAAATGATTTTTTTTTCAGGGGGTAGGGTAGGCAGGAAGCTTAAACTGGGTAATTTAG
TTGTAGAAGAGTGCCCTGTGGCAAATAATTGATTATTCATTTCCAGCATCCCCTTTTCTTCTCCTTGACAGTTATTA
AAAAAAAAAAAGTTACCAGCTTATGTCATTTTAAAGAACACTCGCCCTGAAAACTTCTGAGAGGTTGGCCATTTGAA
ACCCTGGTTTTAGTGTCTGTATTATTAGTGAACTACCGTGTTCCCATGTGGCTACACAACCACAATTATGTACTATC
TGGCTCTTTACCAAAGTTTGCAGACCTCTAATCTAGAGTGCGACATTTCCCCTCATTAACTCTTAGGTCCCTTGGCT
CTAAAAGGGTATATTCATCTTGGCCCTATACAGGGAAAGGGGGAATGGGATTAATGATGTGCTTTGTAAGAAGAACC
AATTTTAATTTTCACAAAGGCTTGACGTAGCTGTGAGAGAAAGGGTAAGAAGAAGCAGGCTTCTTCTTAGAAGTCTG
AGATGGCCTAAAGTGGTGGGGGAAGAAGGGAGAGTGGGGAGAAAGAGAAACAAGAAAAGCTGAGAGTGAATTCCCCA
GAGAGGTAGCCACTGATTCTGCCCTACTCTTTGCTGGAATTCTGGAAAACACCTGGGCTTCTAAAAGATAGGGAGCT
CATGCATCATGGTAGGGCCACAGCTCAGGCTAGGGCCAGAGATAGCTCAGAGTAGCGCCACGGCTCAGGGTAGGTCC
ACAGTGCAGGGTAGGGCCATAGCTCAGAGTAGGGCCATAGCTCATAACCACAGCTCAGGGTAGGACCTGCTGATCTA
TTTGGGGACCCCCAGCAGAGCCTGTCTAATTGCATATCTTGAAAAGGATTGGAAAACTGTCATAATGACATTATTCC
CTCTCACTTTCCTTGTCCAGGAAAGCCCAGCAGAATCGGAGAGGCTTTTCCGAGGAGCAGCTTCGCCAGGGACAGAA
CGTAATAGGCCTGCAGATGGGCAGCAACAAGGGAGCCTCCCAGGCGGGCATGACAGGGTACGGGATGCCCAGGCAGA
TCATGTAGGACGCGGCATCCTGCCCCTGGTAGAGAGGACGAATGTTCCACACCATGGTCTCTACGAAAAAGAAATAG
TTAGTCACCTTCTGACCTTCTCCTCTTTCTCAAAGCCTTCTGTCCCTGGTTTTTGCAAGTGCTGCATTTCCGCCGAG
AATCCGCGTTGCCTACTGCTGCCACCTCCTGTTCATTTAGAACTATGCAAAGACTCCGCTTCCGTTTTCCTGAGCTC
CTCGGGCCCCAGAGTCTCTGTTTGATTATTTATTTATTTATTTATTTATTTGCCAAAAATTCTCCTCTTCAACTTAT
AGAATGCACCTAATAAAGTAATTAGTCTTGTGTCTTACAGTG

SEQ ID NO: 1627
>HUMOSTRO_PEA_1_PEA_1_P21
MRIAVICFCLLGITCAIPVKQADSGSSEEKQLYNKYPDAVATWLNPDPSQKQNLLAPQVFLNFS

SEQ ID NO: 1628
>HUMOSTRO_PEA_1_PEA_1_P25
MRIAVICFCLLGITCAIPVKQADSGSSEEKQH

SEQ ID NO: 1629
>HUMOSTRO_PEA_1_PEA_1_P30
MRIAVICFCLLGITCAIPVKQADSGSSEEKQVSIFYVFI

SEQ ID NO: 1630
RPL19 -amplicon
TGGCAAGAAGAAGGTCTGGTTAGACCCCAATGAGACCAATGAAATCGCCAATGCCAACTCCCGTCAGCAGATCCGGA
AGCTCATCAAAGATGGGCTGATCA

SEQ ID NO: 1631
TATA box Forward primer
CGGTTTGCTGCGGTAATCAT

SEQ ID NO: 1632
TATA box Reverse primer
TTTCTTGCTGCCAGTCTGGAC

SEQ ID NO: 1633
TATA box -amplicon
CGGTTTGCTGCGGTAATCATGAGGATAAGAGAGCCACGAACCACGGCACTGATTTTCAGTTCTGGGAAAATGGTGTG
CACAGGAGCCAAGAGTGAAGAACAGTCCAGACTGGCAGCAAGAAA

SEQ ID NO: 1634
H61775seg8F2
GAAGGCTCTTGTCACTTACTAGCCAT

SEQ ID NO: 1635
H61775seg8R2

TGTCACCATATTTAATCCTCCCAA

SEQ ID NO: 1636
H61775seg8
GAAGGCTCTTGTCACTTACTAGCCATGTGATTTTGGAAAGAAACTTAACATTAATTCCTTCAGCTACAATGGAATTC
TTGGGAGGATTAAATATGGTGACA

SEQ ID NO: 1637
M85491seg24F
GGCGTCTTTCTCCCTCTGAAC

SEQ ID NO: 1638
M85491seg24R
GTCCCATTCTGGGTGCTGTG

SEQ ID NO: 1639
M85491seg24
GGCGTCTTTCTCCCTCTGAACCTCAGTTTCCACCTGTGTCGAGTGTGGGTGAGACCCCTCGCGGGGAGCTATGCAGG
TTACGGAGAAAAGGCAGCACAGCACCCAGAATGGGAC

SEQ ID NO: 1640
Z21368 junc17-21 Forward primer
GGACGGATACAGCAGGAACG

SEQ ID NO: 1641
Z21368 junc17-21 Reverse amplicon
TATTTTCCAAAAAAGGCCAGCTC

SEQ ID NO: 1642
Z21368 junc17-21 Amplicon
GGACGGATACAGCAGGAACGAAAAAAACATCCGACCCAACATTATTCTTGTGCTTACCGATGATCAAGATGTGGAGCT
GGCCTTTTTTGGAAAATA

SEQ ID NO: 1643
Forward primer Z21368seg39F
GTTGCATTTCTCAGTGCTGGTTT

SEQ ID NO: 1644

Reverse primer Z21368seg39R
AGGGTGCCGGGTGAGG

SEQ ID NO: 1645
Amplicon Z21368seg39:
GTTGCATTTCTCAGTGCTGGTTTCTAATCAGACCAGTGGATTGAGTTTCTCTACCATCCTCCCCACGTTCTTCTCTA
AGCTGCCTCCAAGCCTCACCCGGCACCCT

SEQ ID NO: 1646
HUMGRP5Ejunc3-7F
ACCAGCCACCTCAACCCA

SEQ ID NO: 1647
HUMGRP5Ejunc3-7R
CTGGAGCAGAGAGTCTTTGCCT

SEQ ID NO: 1648
HUMGRP5Ejunc3-7
ACCAGCCACCTCAACCCAAGGCCCTGGGCAATCAGCAGCCTTCGTGGGATTCAGAGGATAGCAGCAACTTCAAAGAT
GTAGGTTCAAAAGGCAAAGACTCTCTGCTCCAG

SEQ ID NO: 1649
Z44808junc8-11 Forward primer
GAAGGCACAGGAAAAACAGATATTG

SEQ ID NO: 1650
Z44808junc8-11 Reverse primer
TGGTGCTCTTGGTCACAGGAT

SEQ ID NO: 1651
Z44808junc8-11 Amplicon:
GAAGGCACAGGAAAAACAGATATTGCATCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAA
AACCAATAAGAATTCAGTGTCATCCTGTGACCAAGAGCACCA

SEQ ID NO: 1652

Forward primer AA161187 seg17F2
CCCTGTGCCTTATTTGACCCT

SEQ ID NO: 1653
Reverse primer AA161187 seg17R2
GCTGGGTAGACTGGGTGCA

SEQ ID NO: 1654
Amplicon AA161187 seg25:
CCTGTGCCTTATTTGACCCTCATGCCAACCCCGGGAGGTGGAGACTGTTGCCCCACTCTGCAGATGCAGAAACGGAG
GCTTGGCTGCTGCCAGGGGGAGGA

SEQ ID NO: 1655
Forward primer -M62069 seg19F
GCTGATTGTCCCCATGAAGG

SEQ ID NO: 1656
Reverse primer- M62069 seg19
TGGCATACGGGAACTCAGTG

SEQ ID NO: 1657
Amplicon:
GCTGATTGTCCCCATGAAGGCCAGCCTTGAAGCTTGGTCAGTCTCCCTAACTGTATGATTGATCCCCACTTATTGCA
CTACATCACTGAGTTCCCGTATGC

SEQ ID NO: 1658
Forward primer -M62069 seg29F
ATTGAATAATTCAGCACCTGAGGC

SEQ ID NO: 1659
Reverse primer- M62069 seg29R
TTCATATGGCTACTCCCCACCT

SEQ ID NO: 1660
Amplicon:
ATTGAATAATTCAGCACCTGAGGCTGGTGGATGATTCTTTGCAATTTGGCAGGAATGGGAGAGTCGGGAGCAGTAGT
TGGCAAGGTGGGGAGTAGCCATATGAA

SEQ ID NO: 1661
Forward primer -HUMCA1X1A seg55F: TTCTCATAGTATTCCATTGATTGGGTA

SEQ ID NO: 1662
Reverse primer- HUMCA1X1A seg55R
CACCGGTATGGAGAATAGCGA

SEQ ID NO: 1663
Amplicon:
TTCTCATAGTATTCCATTGATTGGGTATACCAGGTTCTGTTTACTTTTACTTGGCAGTTGATAGAATAGGTGTAGTT
TATACTTTTTCGCTATTCTCCATACCGGTG

SEQ ID NO: 1664
Forward primer:
ACCCCAAACCCAACTTGATTC

SEQ ID NO: 1665
Reverse primer:
TCAGTGGTGGAGCCAAGTCTC

SEQ ID NO: 1666
Amplicon
ACCCCAAACCCAACTTGATTCCTGCCATATGGAGGAGGCTCTGGAGTCCTGCTCTGTGTGGTCCAGGTCCTTTCCAC
CCTGAGACTTGGCTCCACCACTGA

SEQ ID NO: 1667
Forward primer:
CTCCTGAACCCTACTCCAAGCA

SEQ ID NO: 1668
Reverse primer:
CAGGCGATCCTATGGAAATCC

SEQ ID NO: 1669
Amplicon
CTCCTGAACCCTACTCCAAGCACAGCCTCTGTCTGACTCCCTTGTCCTTCAAGAGAACTGTTCTCCAGGTCTCAGGG
CCAGGATTTCCATAGGATCGCCTG

SEQ ID NO: 1670
Forward primer Z25299 seg23F:
CAAGCAATTGAGGGACCAGG

SEQ ID NO: 1671
Reverse primer Z25299 seg23R: CAAAAAACATTGTTAATGAGAGAGATGAC

SEQ ID NO: 1672
Amplicon Z25299 seg23F :
CAAGCAATTGAGGGACCAGGAAGTGGATCCTCTAGAGATGAGGAGGCATTCTGCTGGATGACTTTTAAAAATGTTTT
CTCCAGAGTCATCTCTCTCATTAACAATGTTTTTTG
SEQ ID NO: 1673
HSSTROL3 seg24 Forward Primer:
ATTTCCATCCTCAACTGGCAGA

SEQ ID NO: 1674
HSSTROL3 seg24 Reverse Primer:
TGCCCTGGAACCCACG

SEQ ID NO: 1675
HSSTROL3 seg24 Amplicon:
ATTTCCATCCTCAACTGGCAGAGATGAGAGCCTGGAGCATTGCAGATGCCAGGGACTTCACAAATGAAGGCACAGCA
TGGGAAACCTGCGTGGGTTCCAGGGCA

SEQ ID NO: 1676
HSSTROL3 seg20-21 Forward primer HSSTROL3 seg20-21F: TCTGCTGGCCACTGTGACTG

SEQ ID NO: 1677
HSSTROL3 seg20-21 Reverse primer HSSTROL3 seg20-21R: GAAGAAAAAGAGCTCGCCTCG

SEQ ID NO: 1678
HSSTROL3 seg20-21 Amplicon HSSTROL3 seg20-21:
TCTGCTGGCCACTGTGACTGCAGCATATGCCCTCAGCATGTGTCCCTCTCTCCCACCCCAGCCAGACGCCCCGCCAG
ATGCCTGTGAGGCCTCCTTTGACGCGGTCTCCACCATCCGAGGCGAGCTCTTTTTCTTC

SEQ ID NO: 1679
Forward primer HSSTROL3 junc21-27F: ACATTTGGTTCTTCCAAGGGACTAC

SEQ ID NO: 1680
Reverse primer HSSTROL3 junc21-27R:
TCGATCTCAGAGGGCACCC

SEQ ID NO: 1681
Amplicon HSSTROL3 junc21-27:
ACATTTGGTTCTTCCAAGGGACTACTGGCGTTTCCACCCCAGCACCCGGCGTGTAGACAGTCCCGTGCCCCGCAGGG
CCACTGACTGGAGAGGGGTGCCCTCTGAGATCGA

SEQ ID NO: 1682
R11723seg13F
ACACTAAAAGAACAAACACCTTGCTC

SEQ ID NO: 1683
R11723seg13R
TCCTCAGAAGGCACATGAAAGA

SEQ ID NO: 1684
R11723seg13 - amplicon:
ACACTAAAAGAACAAACACCTTGCTCTTCGAGATGAGACATTTTGCCAAGCAGTTGACCACTTAGTTCTCAAGAAGC
AACTATCTCTTTCATGTGCCTTCTGAGGA

SEQ ID NO: 1685
R11723junc11-18F
AGTGATGGAGCAAAGTGCCG

SEQ ID NO: 1686
R11723 junc11-18R
CAGCAGCTGATGCAAACTGAG

SEQ ID NO: 1687
R11723 junc11-18 - amplicon:
AGTGATGGAGCAAAGTGCCGGGATCATGTACCGCAAGTCCTGTGCATCATCAGCGGCCTGTCTCATCGCCTCTGCCG
GGTACCAGTCCTTCTGCTCCCCAGGGAAACTGAACTCAGTTTGCATCAGCTGCTG

SEQ ID NO: 1688
H53626 junc24-27F1R3 Forward primer:
GTCCTTCCAGTGCAAGACCCA

SEQ ID NO: 1689
H53626 junc24-27F1R3 Reverse primer:
TGGGCCTGGCAAAGCC

SEQ ID NO: 1690
H53626 junc24-27F1R3 Amplicon:
GTCCTTCCAGTGCAAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTCGGCCACTAGCCTGCCGTGGCC
CGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCATCCTGGGCACCCTGCTCCTGTGGCTTTGCCAGGCCCA

SEQ ID NO: 1691
H53626 seg25Forward primer:
CCGACGGCTCCTACCTCAA

SEQ ID NO: 1692
H53626 seg25Reverse primer:
GGAAGCTGTAGCCCATGGTGT

SEQ ID NO: 1693
H53626 seg25Amplicon:
CCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCGTGCCCGCCAGGACGATGCGGGCATGTACATCTGCCTTGGC
GCCAACACCATGGGCTACAGCTTCC

SEQ ID NO:1694
> Q9P2J2
EGLGEQASWAMVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRP
PLHVIEWLRFGFLLPIFIQFGLYSPRIDPDYVGRVRLQKGASLQIEGLRVEDQGWYECRV
FFLDQHIPEDDFANGSWVHLTVNSPPQFQETPPAVLEVQELEPVTLRCVARGSPLPHVTW
KLRGKDLGQGQGQVQVQNGTLRIRRVERGSSGVYTCQASSTEGSATHATQLLVLGPPVIV
VPPKNSTVNASQDVSLACHAEAYPANLTYSWFQDNINVFHISRLQPRVRILVDGSLRLLA
TQPDDAGCYTCVPSNGLLHPPSASAYLTVLYPAQVTAMPPETPLPIGMPGVIRCPVRANP
PLLFVSWTKDGKALQLDKFPGWSQGTEGSLIIALGNEDALGEYSCTPYNSLGTAGPSPVT
RVLLKAPPAFIERPKEEYFQEVGRELLIPCSAQGDPPPVVSWTKVGRGLQGQAQVDSNSS
LILRPLTKEAHGHWECSASNAVARVATSTNVYVLGTSPHVVTNVSVVALPKGANVSWEPG
FDGGYLQRFSVWYTPLAKRPDRMHHDWVSLAVPVGAAHLLVPGLQPHTQYQFSVLAQNKL
GSGPFSEIVLSAPEGLPTTPAAPGLPPTEIPPPLSPPRGLVAVRTPRGVLLHWDPPELVP
KRLDGYVLEGRQGSQGWEVLDPAVAGTETELLVPGLIKDVLYEFRLVAFAGSFVSDPSNT
ANVSTSGLEVYPSRTQLPGLLPQPVLAGVVGGVCFLGVAVLVSILAGCLLNRRRAARRRR

KRLRQDPPLIFSPTGKSAAPSALGSGSPDSVAKLKLQGSPVPSLRQSLLWGDPAGTPSPH
PDPPSSRGPLPLEPICRGPDGRFVMGPTVAAPQERSGREQAEPRTPAQRLARSFDCSSSS
PSGAPQPLCIEDISPVAPPPAAPPSPLPGPGPLLQYLSLPFFREMNVDGDWPPLEEPSPA
APPDYMDTRRCPTSSFLRSPETPPVSPRESLPGAVVGAGATAEPPYTALADWTLRERLLP
GLLPAAPRGSLTSQSSGRGSASFLRPPSTAPSAGGSYLSPAPGDTSSWASGPERWPRREH
VVTVSKRRNTSVDENYEWDSEFPGDMELLETLHLGLASSRLRPEAEPELGVKTPEEGCLL
NTAHVTGPEARCAALREEFLAFRRRRDATRARLPAYRQPVPHPEQATLL

SEQ ID NO:1695
> AAQ88495
MVWCLGLAVLSLVISQGADGRGKPEVVSVVGRAGESVVLGCDLLPPAGRPPLHVIEWLRFGFLLPIFIQF
GLYSPRIDPDYVGRVRLQKGASLQIEGLRVEDQGWYECRVFFLDQHIPEDDFANGSWVHLTVNSPPQFQE
TPPAVLEVQELEPVTLRCVARGSPLPHVTWKLRGKDLGQGQGQVQVQNGTLRIRRVERGSSGVYTCQASS
TEGSATHATQLLVLGPPVIVVPPKNSTVNASQDASLACHAEAYPANLTYSWFQDNINVFHISRLQPRVRI
LVDGSLRLLATQPDDAGCYTCVPSNGLLHPPSASAYLTVLYPAQVTAMPPETPLPIGMPGVIRCPVRANP
PLLFVSWTKDGKALQLDKFPGWSQGTEGSLIIALGNEDALGEYSCTPYNSLGTAGPSPVTRVLLKAPPAF
IERPKEEYFQEVGRELLIPCSAQGDPPPAAPPSPLPGPGPLLQYLSLPFFREMNVDGDWPPLEEPSPAAP
PDYMDTRRCPTSSFLRSPETPPVSPRESLPGAVVGAGATAEPPYTALADWTLRERLLPGLLPAAPRGSLT
SQSSGRGSASFLRPPSTAPSAGGSYLSPAPGDTSSWASGPERWPRREHVVTVSKRRNTSVDENYEWDSEF
PGDMELLETLHLGLASSRLRPEAEPELGVKTPEEGCLLNTAHVTGPEARCAALREEFLAFRRRRDATRAR
LPAYRQPVPHPEQATLL

SEQ ID NO:1696
> Q9BSH7
(SEE VTNC_HUMAN)

SEQ ID NO:1697
> Q7Z2W2
MKYSCCALVLAVLGTELLGSLCSTVRSPRFRGRIQQERKNIRPNIILVLTDDQDVELGSL
QVMNKTRKIMEHGGATFINAFVTTPMCCPSRSSMLTGKYVHNHNVYTNNENCSSPSWQAM
HEPRTFAVYLNNTGYRTVFFGKYLNEYNGSYIPPGWREWLGLIKNSRFYNYTVCRNGIKE
KHGFDYAKDYFTDLITNESINYFKMSKRMYPHRPVMMVISHAAPHGPEDSAPQFSKLYPN
ASQHITPSYNYAPNMDKHWIMQYTGPMLPIHMEFTNILQRKRLQTLMSVDDSVERLYNML
VETGELENTYIIYTADHGYHIGQFGLVKGKSMPYDFDIRVPFFIRGPSVEPGSIVPQIVL
NIDLAPTILDIAGLDTPPDVDGKSVLKLLDPEKPGNRFRTNKKAKIWRDTFLVERGKFLR
KKEESSKNIQQSNHLPKYERVKELCQQARYQTACEQPGQKWQCIEDTSGKLRIHKCKGPS
DLLTVRQSTRNLYARGFHDKDKECSCRESGYRASRSQRKSQRQFLRNQGTPKYKPRFVHT

RQTRSLSVEFEGEIYDINLEEEEELQVLQPRNIAKRHDEGHKGPRDLQASSGGNRGRMLA
DSSNAVGPPTTVRVTHKCFILPNDSIHCERELYQSARAWKDHKAYIDKEIEALQDKIKNL
REVRGHLKRRKPEECSCSKQSYYNKEKGVKKQEKLKSHLHPFKEAAQEVDSKLQLFKENN
RRRKKERKEKRRQRKGEECSLPGLTCFTHDNNHWQTAPFWNLGSFCACTSSNNNTYWCLRTVNETHNFLFCEFATGF
LEYFDMNTDPYQLTNTVHTVERGILNQLHVQLMELRSCQGYKQ
CNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG

SEQ ID NO:1698
> AAH12997
LRSCQGYKQCNPRPKNLDVGNKDGGSYDLHRGQLWDGWEG

SEQ ID NO:1699
> Q8N441
MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPPL
TMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVVLDDI
SPGKESLGPDSSSGGQEDPASQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVASGHPRP
DITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQ
RTRSKPVLTGTHPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGRHNSTIDVGG
QKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRSAFLTVLPDPK
PPGPPVASSSSATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPP
GTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKLYPKLYTDIHT
HTHTHSHTHSHVEGKVHQHIHYQC

SEQ ID NO:1700
> Q9H4D7
MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPPL
TMWTKDGRTIHSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVVLDDI
SPGKESLGPDSSSGGQEDPASQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVASGHPRP
DITWMKDDQALTRPEAAEPRKKKWTLSLKNLRPEDSGKYTCRVSNRAGAINATYKVDVIQ
RTRSKPVLTGTHPVNTTVDFGGTTSFQCKVRSDVKPVIQWLKRVEYGAEGRHNSTIDVGG
QKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRSAFLTVLPDPK
PQGPPVASSSSATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPP
GTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKLYPKLYTDIHT
HTHTHSHTHSHVEGKVHQHIHYQC

SEQ ID NO:1701:

> Q9HAP5

MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPANLQLCHGIEYQNMRL
PNLLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFAPVCLDDLDETIQPCHSLCVQ
VKDRCAPVMSAFGFPWPDMLECDRFPQDNDLCIPLASSDHLLPATEEAPKVCEACKNKND
DDNDIMETLCKNDFALKIKVKEITYINRDTKIILETKSKTIYKLNGVSERDLKKSVLWLK
DSLQCTCEEMNDINAPYLVMGQKQGGELVITSVKRWQKGQREFKRISRSIRKLQC

SEQ ID NO:1702

> AAA59968 (see DCOR_HUMAN)

SEQ ID NO:1703

> AAH14562

MQLKCNDSKAIVKTLAATGTGFDCASKTEIQLVQSLGVPPERIIYANPCKQVSQIKYAAN
NGVQMMTFDSEVELMKVARAHPKAKLVLRIATDDSKAVCRLSVKFGATLRTSRLLLERAK
ELNIDVVGVSFHVGSGCTDPETFVQAISDARCVFDMGAEVGFSMYLLDIGGGFPGSEDVK
LKFEEITGVINPALDKYFPSDSGVRIIAEPGRYYVASAFTLAVNIIAKKIVLKEQTGSDD
EDESSEQTFMYYVNDGVYGSFNCILYDHAHVKPLLQKRPKPDEKYYSSSIWGPTCDGLDR
IVERCDLPEMHVGDWMLFENMGAYTVAAASTFNGFQRPTIYYVMSGPAWQLMQQFQNPDF
PPEVEEQDASTLPVSCAWESGMKRHRAACASASINV

SEQ ID NO:1704

> Q9NWT9

MRPRSGPTRNPRLRAFAGVPTRGRTRGQSRRCAAEASAGPERDARPGAPAAGTMGAAHSASEEVRELEGKTGFSSDQ
IEQLHRRFKQLSGDQPTIRKENFNNVPDLELNPIRSKIVRAFF
DNRNLRKGPSGLADEINFEDFLTIMSYFRPIDTTMDEEQVELSRKEKLRFLFHMYDSDSD
GRITLEEYRNVKWSRSCCRETLTSRRSPLAPSPTGP

SEQ ID NO:1705

> Q8IXD7

MRILQLILLALATGLVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGATLIAPRWLLTA
AHCLKPWVSLTSPTHVSPDLSSSNYCLSHLSRYIVHLGQHNLQKEEGCEQTRTATESFPH
PGFNNSLPNKDHRNDIMLVKMASPVSITWAVRPLTLSSRCVTAGTSCLISGWGSTSSPQL
RLPHTLRCANITIIEHQKCENAYPGNITDTMVCASVQEGGKDSCQGDSGGPLVCNQSLQG
IISWGQDPCAITRKPGVYTKVCKYVDWIQETMKNN

SEQ ID NO:1706

> Q9NS21

MSLLPRRAPPVSMRLLAAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKY

PHCEEKMVIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNAWNEKRRFYEE

SEQ ID NO:1707

> Q8IXM0

MYAQALLVVGVLQRQAAAQHLHEHPPKLLRGHRVQERVDDRAEVEKRLREGEEDHVRPEV

GPRPVVLGFGRSHDPPNLVGHPAYGQCHNNQPWADTSRRERQRKEKHSMRTQ

SEQ ID NO:1708

> Q96AC2

MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVME

QSAGIMYRKSCASSAACLIASAGYQSFCSPGKLNSVCISCCNTPLCNGPRPKKRGSSASA

LRPGLRTTILFLKLALFSAHC

SEQ ID NO:1709

> Q8N2G4

MWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDCSSPEFIVNCTVNVQDMCQKEVME

QSAGIMYRKSCASSAACLIASAGYQSFCSPGKLNSVCISCCNTPLCNGPRPKKRGSSASA

LRPGLRTTILFLKLASSRHTAELKEMPPPPALFFQPSPPTPHLPE

SEQ ID NO:1710

> BAC85518

MQAPRAAPAAPLSYDRRPRDSGRMWVLGIAATFCGLFLLPGFALQIQCYQCEEFQLNNDC

SSPEFIVNCTVNVQDMCQKEVMEQSAGIMYRKSCASSAACLIASAGYQSFCSPGKLNSVC

ISCCNTPLCSGPRPKKRGSSASALRPGLRTTILFLRLALFSAHC